

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:57 ; Search time 18 seconds
(without alignments)
670.578 Million cell updates/sec

Title: US-09-466-778-11

Perfect score: 1857

Sequence: 1 MTGPGKHCKECSHYVVDGL.....ALAAYSYFRINRKTIGFXHF 353

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:*

1: /SIDSL1/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SIDSL1/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SIDSL1/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SIDSL1/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SIDSL1/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SIDSL1/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SIDSL1/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SIDSL1/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SIDSL1/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SIDSL1/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SIDSL1/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SIDSL1/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SIDSL1/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SIDSL1/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SIDSL1/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SIDSL1/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SIDSL1/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SIDSL1/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SIDSL1/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SIDSL1/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SIDSL1/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262	14.1	277	13 R26044	Tumour necrosis fa
2	262	14.1	277	18 W13654	Tumour necrosis fa
3	262	14.1	277	20 W4087	Tumour necrosis fa
4	187.5	10.1	360	20 Y1381	Amino acid sequenc
5	169.5	9.1	339	21 Y57083	Rat proteoglycan 1
6	165.5	8.9	354	21 Y57080	Human link protein
7	165.5	8.9	355	21 Y57082	Chicken proteoglyc
8	163.5	8.8	354	21 Y57081	Human proteoglycan
9	163.5	8.8	354	21 Y57084	Human proteoglycan
10	157.5	8.5	1257	15 R46627	Neurocan core prot
11	156.5	8.4	252	19 W74523	Cloned duplicate o
12	155.5	8.4	355	16 R77034	Rat brain-enriched

13	155.5	8.4	912	16 R85442	Bovine brevic acid
14	155	8.3	528	19 W59882	Amino acid sequenc
15	151	8.1	116	19 W74522	Amino acid sequenc
16	150.5	8.1	2409	12 R12609	Version. Homo sa
17	148	8.0	908	16 R85443	Rat brevic acid
18	147	7.9	378	16 R77035	Cat brain-enriched
19	144	7.8	354	15 R57350	Murine delta prote
20	114.5	6.2	722	21 Y79028	Human cartilage li
21	109	5.9	322	20 Y13379	Amino acid sequenc
22	109	5.9	322	21 Y87287	Human signal pepti
23	107	5.8	322	19 W56249	Amino acid sequenc
24	107	5.8	2189	11 R05222	Antigen GX5401FL e
25	105	5.7	810	18 W37500	Human nel-related
26	99.5	5.4	727	18 W11719	C-Delta-1 polypept
27	99.5	5.4	740	18 W00876	C-Delta-1 polypept
28	98.5	5.3	728	21 Y79029	Chick delta protei
29	97.5	5.3	503	12 R14768	Metastasis-specifi
30	97	5.2	116	20 Y12323	Human 5' EST seque
31	96	5.2	361	13 R20816	Haematopoietic CD4
32	96	5.2	361	17 R91444	Human haematopoiet
33	96	5.2	361	19 W80453	Human CD44 antigen
34	96	5.2	361	20 W86200	Human CD44 antigen
35	96	5.2	493	13 R20817	Epithelial CD44 An
36	96	5.2	493	17 R91445	Human epithelial C
37	96	5.2	493	19 W80454	Human CD44 antigen
38	96	5.2	493	20 W89151	Human CD44 antigen
39	95.5	5.1	1480	13 R25079	Drosophila SLIT pr
40	94.5	5.1	722	18 W11720	M-Delta-1 polypept
41	93.5	5.0	512	20 W75494	Truncated human de
42	93.5	5.0	520	18 W18348	Proliferation and
43	93.5	5.0	702	18 W18349	Proliferation and
44	93.5	5.0	702	20 W75495	Truncated human de
45	93.5	5.0	723	18 W18353	Proliferation and

ALIGNMENTS

RESULT 1	
R26044	
ID	R26044 standard; Protein; 277 AA.
XX	
AC	R26044;
XX	
XX	
DT	03-FEB-1993 (first entry)
XX	
DE	Tumour necrosis factor-induced glycoprotein TSG-6.
XX	
XX	TNF; chronic inflammatory conditions; rheumatoid arthritis; sepsis;
KW	cancer; infections.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key Location/Qualifiers
FT	Peptide 1..19
FT	/*tag= a
FT	/note= "putative signal sequence"
FT	41..44
FT	/*tag= b
FT	/note= "potential N-glycosylation site"
FT	118..120
FT	/*tag= c
FT	/note= "potential N-glycosylation site"
FT	256..260
FT	/*tag= d
FT	/note= "potential N-glycosylation site"
XX	
PN	W09212175-A.
XX	
PD	23-JUL-1992.
XX	
PF	14-JAN-1992; 92WO-US00333.
XX	

SQ Sequence 360 AA;

Query Match 10.1%; Score 187.5; DB 20; Length 360;
 Best Local Similarity 30.8%; Pred. No. 1.8e-11;
 Matches 45; Conservative 24; Mismatches 48; Indels 29; Gaps 5;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHLRSPL 64
 Db 142 gryrce-----vidglede-----sglvelelrgv---vfpqspn 174
 QY 65 GOYKLFDRKAREACANEATMATYNQLSYQKAKYHLCSSAGWLETGRVAYPTAFASQNCG 124
 Db 175 gryqfifhegqqvcaqaavvasfedlfraweegldwcnagwldqatvqyplmrpqcg 234

QY 125 S-GVV-GIVDYGPRPNKSEMDVFCY 148
 Db 235 gpglapgvrsygprrhlrlhrydvfcf 260

RESULT 5
 Y57083
 ID Y57083 standard; protein; 339 AA.
 AC Y57083;
 DT 28-FEB-2000 (first entry)
 XX XX
 DE Rat proteoglycan link protein precursor.
 XX Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;
 KW uveoretinitis; inflammatory response.
 XX Rattus sp.
 OS WO9956763-AL.
 PN 11-NOV-1999.
 PD 07-MAY-1999; 99WO-US10250.
 PF 07-MAY-1998; 98US-0084636.
 PR (REGC) UNIV CALIFORNIA.
 PA Kaufman DL, Tian J, Olcott A;
 XX WPI; 2000-052905/04.
 DR Administration of neglected target tissue antigens to modulate immune
 PT responses -
 XX Disclosure; Page 27; 79pp; English.

XX Amino acid sequences Y57063-Y57091 are examples of neglected target
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)
 CC not involved in autoimmunity. These peptides and proteins are used in the
 CC method of the invention which involves administering an NNTA as an
 CC antigen based immunotherapeutic agent, to a host afflicted with an
 CC autoimmune response associated with an autoimmune disease. The
 CC immunotherapeutic agent is used to treat autoimmune diseases such as
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory
 CC immune responses. The NNTA induces regulatory tolerance by elicitation of
 CC regulatory T cells among T cells recognizing the NNTA but not
 CC participating in the immune response. The NNTA are capable of recognition
 CC by substantial populations of uncommitted T cells which can be primed, or
 CC biased, towards regulatory responses to provide effective treatment. The
 CC NNTA target determinants used as agents promoting tolerance agents have failed
 CC to induce an effective regulatory T cell response. NNTAs as agents

CC promoting tolerance are anticipated to be safer than use of target
 CC determinants.
 XX
 SQ Sequence 339 AA;
 Query Match 9.1%; Score 169.5; DB 21; Length 339;
 Best Local Similarity 30.8%; Pred. No. 1.3e-09;
 Matches 45; Conservative 20; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHLRSPL 64
 Db 120 grykce-----viegl-----edttavvalelqgv---vfpfprl 152
 QY 65 GOYKLFDRKAREACANEATMATYNQLSYQKAKYHLCSSAGWLETGRVAYPTAFASQNCG 124
 Db 153 grynlfnhearqacldqdviasfdqldawrgldwcnagwldsgvqyptkprepg 212

QY 125 --SGVVGIVDYGPRPNKSEMDVFCY 148
 Db 213 gqntvpgvrnygfwdkksrydvfcf 238

RESULT 6
 Y57080
 ID Y57080 standard; protein; 354 AA.
 AC Y57080;
 DT 28-FEB-2000 (first entry)
 XX XX
 DE Human link protein precursor amino acid sequence.
 XX Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;
 KW uveoretinitis; inflammatory response.
 XX Homo sapiens.
 OS WO9956763-AL.
 PN 11-NOV-1999.
 PD 07-MAY-1999; 99WO-US10250.
 PF 07-MAY-1998; 98US-0084636.
 PR (REGC) UNIV CALIFORNIA.
 PA Kaufman DL, Tian J, Olcott A;
 XX WPI; 2000-052905/04.
 DR Administration of neglected target tissue antigens to modulate immune
 PT responses -
 XX Disclosure; Page 26; 79pp; English.

XX Amino acid sequences Y57063-Y57091 are examples of neglected target
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)
 CC not involved in autoimmunity. These peptides and proteins are used in the
 CC method of the invention which involves administering an NNTA as an
 CC antigen based immunotherapeutic agent, to a host afflicted with an
 CC autoimmune response associated with an autoimmune disease. The
 CC immunotherapeutic agent is used to treat autoimmune diseases such as
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory
 CC immune responses. The NNTA induces regulatory tolerance by elicitation of
 CC regulatory T cells among T cells recognizing the NNTA but not
 CC participating in the immune response. The NNTA are capable of recognition
 CC by substantial populations of uncommitted T cells which can be primed, or
 CC biased, towards regulatory responses to provide effective treatment. The
 CC NNTA target determinants used as agents promoting tolerance agents have failed
 CC to induce an effective regulatory T cell response. NNTAs as agents

CC NNTA are effective in regulating undesirable immune responses even when
 CC target determinants used as agents promoting tolerance agents have failed
 CC to induce an effective regulatory T cell response. NNTAs as agents
 CC promoting tolerance are anticipated to be safer than use of target
 CC determinants.

XX Sequence 354 AA;

Query Match 8.9%; Score 165.5; DB 21; Length 354;
 Best Local Similarity 30.1%; Pred. No. 3.8e-09;
 Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;
 QY 5 GKHKCECKSHYVGDGLNCEPEQLPDRCLQDNGOCHADAKCVDLHFQDTTGVVPHLRSPL 64
 Db 135 grykce-----vieg|-----eddtavvaldlqgv---vfpypfprl 167
 QY 65 GYKLTDFDKAREACANEATMATYNOLSYXOKAKYHLCISAGWLETGRVAYPTAFASQNC 124
 Db 168 grynlnfheaqqacldqdasfdqlydawrsigldwcnagwlsdgsqvypitkprepcg 227
 QY 125 --SGVVGIIVDYGPRNKSEMDVFCY 148
 Db 228 gqntvpgvrnygfwdkdkrsydvfcf 253

RESULT 7
 Y57082
 ID Y57082 standard; protein; 355 AA.
 AC Y57082;
 DT 28-FEB-2000 (first entry)
 XX
 DE Chicken proteoglycan link protein precursor.

XX Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;
 KW uveoretinitis; inflammatory response.

OS Gallus sp.
 PN WO9956763-Al.
 PD 11-NOV-1999.

XX 07-MAY-1999; 99WO-US10250.
 XX 07-MAY-1998; 98US-0084636.

XX (REGC) UNIV CALIFORNIA.
 XX Kaufman DL, Tian J, Olcott A;
 XX WPI; 2000-052905/04.

XX Administration of neglected target tissue antigens to modulate immune
 PT responses -
 XX Disclosure; Page 27; 79pp; English.

XX Amino acid sequences Y57063-Y57091 are examples of neglected target
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)
 CC not involved in autoimmunity. These peptides and proteins are used in the
 CC method of the invention which involves administering an NNTA as an
 CC antigen based immunotherapeutic agent, to a host afflicted with an
 CC autoimmune response associated with an autoimmune disease. The
 CC immunotherapeutic agent is used to treat autoimmune diseases such as
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory
 CC immune responses. The NNTA induces regulatory tolerance by elicitation of
 CC regulatory T cells among T cells recognizing the NNTA but not

CC participating in the immune response. The NNTA are capable of recognition
 CC by substantial populations of uncommitted T cells which can be primed, or
 CC biased, towards regulatory responses to provide effective treatment. The
 CC NNTA are effective in regulating undesirable immune responses even when
 CC target determinants used as agents promoting tolerance agents have failed
 CC to induce an effective regulatory T cell response. NNTAs as agents
 CC promoting tolerance are anticipated to be safer than use of target
 CC determinants.

XX Sequence 355 AA;

Query Match 8.9%; Score 165.5; DB 21; Length 355;
 Best Local Similarity 29.9%; Pred. No. 3.8e-09;
 Matches 44; Conservative 26; Mismatches 46; Indels 31; Gaps 5;
 QY 5 GKHKCECKSHYVGDGLNCEPEQLPDRCLQDNGOCHADAKCVDLHFQDTTGVVPHLRSPL 63
 Db 136 grykce-----vieg|-----eddtavvalnle---gvvfpyspr 167
 QY 64 LGQYKLTDFDKAREACANEATMATYNOLSYXOKAKYHLCISAGWLETGRVAYPTAFASQNC 123
 Db 168 lgrynlnfheaqqacldqdasfdqlydawrsigldwcnagwlsdgsqvypitkprepc 227
 QY 124 G--SGVVGIIVDYGPRNKSEMDVFCY 148
 Db 228 gqntvpgvrnygfwdkdkrsydvfcf 254

RESULT 8
 Y57081
 ID Y57081 standard; protein; 354 AA.
 AC Y57081;
 DT 28-FEB-2000 (first entry)

XX Human proteoglycan link protein precursor amino acid sequence.
 DE Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;
 KW uveoretinitis; inflammatory response.

OS Homo sapiens.
 PN WO9956763-Al.
 PD 11-NOV-1999.

XX 07-MAY-1999; 99WO-US10250.
 XX 07-MAY-1998; 98US-0084636.

XX (REGC) UNIV CALIFORNIA.
 XX Kaufman DL, Tian J, Olcott A;
 XX WPI; 2000-052905/04.

XX Administration of neglected target tissue antigens to modulate immune
 PT responses -
 XX Disclosure; Page 26-27; 79pp; English.

XX Amino acid sequences Y57063-Y57091 are examples of neglected target
 CC tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)
 CC not involved in autoimmunity. These peptides and proteins are used in the
 CC method of the invention which involves administering an NNTA as an
 CC antigen based immunotherapeutic agent, to a host afflicted with an
 CC autoimmune response associated with an autoimmune disease. The
 CC immunotherapeutic agent is used to treat autoimmune diseases such as
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune

FT Modified-site /note= "O-linked glycosylation site"
 FT 436
 FT Modified-site /note= "O-linked glycosylation site"
 FT 445
 FT Modified-site /note= "O-linked glycosylation site"
 FT 454
 FT Modified-site /note= "O-linked glycosylation site"
 FT 456
 FT Modified-site /note= "O-linked glycosylation site"
 FT 511
 FT Modified-site /note= "O-linked glycosylation site"
 FT 572
 FT Modified-site /note= "O-linked glycosylation site"
 FT 576
 FT Modified-site /note= "O-linked glycosylation site"
 FT 579
 FT Modified-site /note= "O-linked glycosylation site"
 FT 655
 FT Modified-site /note= "N-linked glycosylation site"
 FT 689
 FT Modified-site /note= "O-linked glycosylation site"
 FT 7..737
 FT Modified-site /note= "N-linked glycosylation site"
 FT 764
 FT Modified-site /note= "O-linked glycosylation site"
 FT 793
 FT Modified-site /note= "O-linked glycosylation site"
 FT 807
 FT Modified-site /note= "O-linked glycosylation site"
 FT 829
 FT Modified-site /note= "O-linked glycosylation site"
 FT 857
 FT Modified-site /note= "O-linked glycosylation site"
 FT 859
 FT Modified-site /note= "O-linked glycosylation site"
 FT 861
 FT Modified-site /note= "O-linked glycosylation site"
 FT 863
 FT Modified-site /note= "O-linked glycosylation site"
 FT 905
 FT Modified-site /note= "O-linked glycosylation site"
 FT 913
 FT Peptide /note= "O-linked glycosylation site"
 FT 933..942
 FT Binding-site /note= "Used for primer design"
 FT 944..945
 FT Modified-site /note= "Chondroitin sulphate attachment site"
 FT 950
 FT Modified-site /note= "O-linked glycosylation site"
 FT 967
 FT Modified-site /note= "N-linked glycosylation site"
 FT 1136..1147
 FT Peptide /note= "Used in primer design"
 FT 1164
 FT Modified-site /note= "N-linked glycosylation site"
 FT XX
 PN W09403601-A.
 PN 17-FEB-1994.
 PD 03-AUG-1993; 93WO-US07306.
 XX 03-AUG-1992; 92US-0922911.
 XX (UYNV) UNIV NEW YORK STATE.
 PA Margolis RK, Margolis RU, Rauch U;
 PI WPI; 1994-065690/08.
 DR N-PSDB; Q57710.
 XX Eukaryotic neurocan polypeptide(s) with epidermal growth factor,
 PT lectin or complement binding activity - used in the diagnosis,
 PT

PT treatment or research of hypersensitivity and allergic diseases
 XX
 PS Claim 1; Page 69-75; 105pp; English.
 XX
 CC This sequence represents a neurocan polypeptide. This protein has
 CC several biological activities, including cell adhesion, leukocyte-
 CC endothelial cell recognition, tissue-related inflammation allergies,
 CC cellular and/or humoral hypersensitivity, trauma, neuronal
 CC development, and cell transport and/or infection. Compositions
 CC containing them can be used as modulators of these conditions, and
 CC may be used as therapeutic, diagnostic, and/or research tools.
 CC Neurocan peptides can be used to mimic proteins, such as lectins,
 CC cell adhesion molecules, versicans, aggrecans or gelsolins, as
 CC receptor or effector subtypes. The protein can be used to treat
 CC diseases involving a qualitative or quantitative pathological
 CC abnormality of cell adhesion or leukocyte-endothelial cell recognition,
 CC or a functionally associated molecule such as a membrane cytoplasmic
 CC protein, lipid, carbohydrate, saccharide, nucleoside, enzyme or ion.
 XX
 SQ Sequence 1257 AA;
 Query Match 8.5%; Score 157.5; DB 15; Length 1257;
 Best Local Similarity 34.2%; Pred. No. 1.8e-07;
 Matches 41; Conservative 19; Mismatches 53; Indels 7; Gaps 3;
 QY 47 DLHFQDTTGVFHLRSPGLGOYKLTFTDKAREACANEAATMAYNQLSYXQKAKYHLCAGW 106
 DB 150 dlvtlevtgvvfhyraardryaltfaaeqaachlssatlaaqrhlqaafedgfdncdagw 209
 QY 107 LETGRVAYPTAFASQNC---GSGVGVIVDYGPRPNKSEMDVFCYRMKDVNCTKKVGTVG 163
 DB 210 lsdrtrvypitgrpcygydrslpvrsg-rrdpqelydyvcfare---lggevfyvg 265
 RESULT 11
 W74523
 ID W74523 standard; Protein; 252 AA.
 XX
 AC W74523;
 XX
 DT 02-DEC-1998 (first entry)
 XX
 DE Cloned duplicate of human aggrecan GI-B domain.
 XX
 KW Human; aggrecan GI-B domain; homogeneous polysaccharide;
 KW heterogeneous polysaccharide.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 22 /note= "encoded by TCC"
 FT Misc-difference 133 /note= "encoded by TCC"
 FT Misc-difference 240 /note= "encoded by ACT"
 FT Misc-difference 241 /note= "encoded by GCA"
 XX
 PN EP861903-A1.
 XX
 PD 02-SEP-1998.
 XX
 PF 22-FEB-1997; 97EP-0102951.
 XX
 PR 22-FEB-1997; 97EP-0102951.
 XX
 PA (LANS/) LANSING M.
 PA (SCHM/) SCHMIDT G.
 PA (UHLE/) UHLENKUEKEN J.
 XX
 PI Lansing M;

XX WPI; 1998-449114/39.
 DR N-PSDB; V53720; V53721.

XX Production of homogeneous polysaccharides from heterogeneous
 PT polysaccharides used for diagnosis and therapy of diseases -
 FT comprises immobilisation on support e.g. polymer matrix and
 PT selective cleavage with e.g. glucosidase or hydrolase

XX Example 1; Fig 4; 19pp; English.

XX This is the amino acid sequence of the cloned duplicate of human
 CC aggrecan G1-B domain, starting from the RBS of the tet o/p of
 CC pr-HA-20, used in the method of the invention which involves the
 CC production of homogeneous polysaccharides from heterogeneous
 CC polysaccharides. The oligonucleotides and polysaccharides are
 CC useful for the diagnosis and therapy of diseases.

XX Sequence 252 AA;

Query Match 8.4%; Score 156.5; DB 19; Length 252;
 Best Local Similarity 25.9%; Pred. No. 2e-08;
 Matches 49; Conservative 29; Mismatches 96; Indels 15; Gaps 4;

QY 57 VFHLRSLGQYKLTDFDKAREACANEATMATYNQLSYXQKAKYHLCGAGHLETRVAYPT 116
 Db 31 vfhyraistrtyldfdragraciqnsaliatpeqlqaayedghqcdagwladqtvrypi 90

QY 117 AFASQNC---GSGVGVGIVDYGPRPNKSEMDVFCY--RMKDVNCTKKVGVGDFGSYSGN 171
 Db 91 htpregcygdkdefpgvrtgyir-dtnetydvycfaeeigtrgetlevvkvgivfhyrai 149

QY 172 LLQVLMFSPSLNPLTEVLAYSNSAGRAFLFH-----LTDLSIRGTFLFXPQNSG 222
 Db 150 strtyldfdragraciqnsaliatpeqlqaayedghqcdagwladqtvrypihtpregc 209

QY 223 LGENETLSG 231

Db 210 ygdkdefpg 218

RESULT 12
 ID R77034 standard; Protein; 355 AA.
 XX R77034;

XX 28-APR-1996 (first entry)

XX Rat brain-enriched hyaluronan binding protein.

XX Brain-enriched hyaluronan binding protein; BEHAB; glioma;
 KW brain tumour; hyaluronic acid; proteoglycan; tumour marker;
 KW diagnostic.

XX Rattus sp.

XX Key Location/Qualifiers
 FT Peptide 1..22

FT Domain /label= Sig_peptide

FT Domain /label= Ig_fold_domain

FT Domain 158..257

FT Domain /label= PTR1

FT Domain 258..355

FT Domain /label= PTR2

XX WO9527785-A1.

XX 19-OCT-1995.

XX 07-APR-1995; 95WO-US04353.

XX 08-APR-1994; 94US-0225477.
 XX (UYVA) UNIV YALE.
 PA Hockfield S, Jaworski DM;

XX WPI; 1995-366390/47.

DR N-PSDB; T01475.

XX Mammalian brain-enriched hyaluronan-binding protein and its DNA -

PT useful as diagnostic marker for detection of brain tumours and other

PT neuro: pathological states

XX Example 2; Fig 1; 46pp; English.

XX Novel rat brain-enriched hyaluronan binding (BEHAB) protein (R77034)

CC is isolated from brain tissue and has a high degree of homology to

CC members of the proteoglycan tandem repeat family of hyaluronan

CC binding proteins. It includes 2 domains, PTR1 and PTR2, thought

CC to be involved in hyaluronan binding. BEHAB is expressed at markedly

CC increased levels in glioma tissue, and can therefore be used as a

CC diagnostic marker. Recombinant BEHAB is obtd. by expression of the

CC isolated encoding cDNA sequence (T01475) in procaryotic or eucaryotic

CC host cells.

XX Sequence 355 AA;

Query Match 8.4%; Score 155.5; DB 16; Length 355;
 Best Local Similarity 27.9%; Pred. No. 4.3e-08;
 Matches 46; Conservative 21; Mismatches 61; Indels 37; Gaps 5;

QY 10 ECKSHYVDGLNCEPEQLPIDRCLQD-----NGOCHADAKCVDLHF 50
 Db 201 qcdagwlsdqtrvypiqnpreacygmdgpgvrvnygvvgpddlydvycyae----dlN- 255

QY 51 QDTTGVFHLRSLGQYKLTDFDKAREACANEATMATYNQLSYXQKAKYHLCGAGHLETG 110
 Db 256 -----gelflgappg--kitweeardyclergaqiastgqlyaaawnggldrcspgwiadg 308

QY 111 RVAYPTAFASQNGSGVGVGIVDYGPRPN-----KSEMDVFCYR 149
 Db 309 svrypilitpsqrcggcgpgvktlflfpnqtgfpksqnrnfvyvfr 353

RESULT 13

ID R85442 standard; Protein; 912 AA.

XX R85442;

XX 19-FEB-1996 (first entry)

XX Bovine brevican core protein.

XX Brevican; chondroitin sulphate proteoglycan; glial cell; axon;

XX neurofibromatosis; gliosis.

XX Bos taurus.

XX Key Location/Qualifiers

FT Peptide 1..34

FT Region /label= Sig_peptide

FT Region 35..158

FT Modified-site /label= Immunoglobulin-like_loop

FT Modified-site 130

FT Region /label= N-glycosylation_site

FT Modified-site 159..353

FT Modified-site 337

FT Region /label= Protein-like_tandem_repeats

FT Region /label= N-glycosylation_site

FT Region 354..648


```
PR 22-FEB-1997; 97EP-0102951.
XX
XX (LANS/) LANSING M.
PA (SCHM/) SCHMIDT G.
PA (UHLE/) UHLENKUEKEN J.
XX
XX LANSING M;
PI
XX
XX WPI; 1998-449114/39.
DR N-PSDB; V53715.
XX
XX Production of homogeneous polysaccharides from heterogeneous
PT polysaccharides used for diagnosis and therapy of diseases -
PT comprises immobilisation on support e.g. polymer matrix and
PT selective cleavage with e.g. glucosidase or hydrolase
XX
XX Example 1; Fig 3; 19pp; English.
PS
XX This is the amino acid sequence of the human aggrecan G1-B domain,
CC used in the method of the invention which involves the production
CC of homogeneous polysaccharides from heterogeneous polysaccharides.
CC The oligonucleotides and polysaccharides are useful for the diagnosis
CC and therapy of diseases.
XX
XX Sequence 116 AA;
SQ

Query Match 8.1%; Score 151; DB 19; Length 116;
Best Local Similarity 35.8%; Pred. No. 2.4e-08;
Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps 2;

QY 57 VHLRSLGQYKLTEDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAYPT 116
Db 14 vfhyraistrtytdfdraqraciqmsailatpeqlqaayedghqcdagwladqtvypl 73

QY 117 AFAQSQC---GSGVGVIVDYGPRPNKSEMMWDFCY 148
Db 74 htpregcygdkdefpgvrtgyir-dtnetydvycf 107
```

Search completed: April 4, 2001, 13:05:01
Job time: 64 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2001, 13:04:01 ; Search time 27.49 Seconds
(without alignments)
1505.071 Million cell updates/sec
Title: US-09-466-778-11
Perfect score: 1857
Sequence: 1 MTGPGKHCKECKSHYVGDGL.....ALAAYSYFIRNKTTIGFXHF 353
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_15:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1757	94.6	897	4 Q9RY3	Q9ry3 homo sapien
2	1468	79.1	1069	4 Q9UF98	Q9uf98 homo sapien
3	662	35.6	2212	4 Q33072	Q33072 homo sapien
4	645.5	34.8	2570	4 Q9NY15	Q9ny15 homo sapien
5	264	14.2	275	11 O08859	O08859 mus musculus
6	177	9.5	2109	13 P79787	P79787 gallus gall
7	169.5	9.1	355	11 Q921X7	Q921X7 mus musculus
8	162	8.7	1321	4 O4594	O4594 homo sapien
9	161.5	8.7	655	11 O88564	O88564 rattus norv
10	161.5	8.7	1290	13 Q9W6E1	Q9w6e1 gallus gall
11	160.5	8.6	2394	6 O77610	O77610 bos taurus
12	153	8.2	192	6 O02817	O02817 oryctolagus
13	153	8.2	656	6 O77612	O77612 bos taurus
14	151.5	8.2	1643	6 O77611	O77611 bos taurus
15	151.5	8.2	3381	6 O77609	O77609 bos taurus
16	126.5	6.8	103	6 Q9TBT3	Q9tbt3 sus scrofa
17	120.5	6.5	103	6 O46380	O46380 oryctolagus
18	110	5.9	396	13 Q9W6S4	Q9w6s4 gallus gall
19	109	5.9	302	4 Q9UNF4	Q9unf4 homo sapien

20	109	5.9	322	4 Q9F5Y7	Q9f5y7 homo sapien
21	105	5.7	1584	5 Q93791	Q93791 caenorhabdi
22	104	5.6	3507	5 Q23587	Q23587 caenorhabdi
23	103.5	5.6	717	13 P87357	P87357 brachydanio
24	102.5	5.5	778	13 Q9IBG4	Q9ibg4 xenopus lae
25	100	5.4	816	11 O70474	O70474 rattus norv
26	99.5	5.4	682	4 Q9NTM2	Q9ntm2 homo sapien
27	99.5	5.4	728	13 Q90556	Q90556 gallus gall
28	99.5	5.4	802	13 O57462	O57462 brachydanio
29	99.5	5.4	3623	4 O60494	O60494 homo sapien
30	99	5.3	780	11 Q9QX8	Q9qx8 mus musculu
31	99	5.3	1440	5 Q20204	Q20204 caenorhabdi
32	99	5.3	2906	11 Q9WUH9	Q9wuh9 rattus norv
33	98	5.3	294	4 Q92493	Q92493 homo sapien
34	98	5.3	2531	5 O16004	O16004 lytechinus
35	97.5	5.3	780	11 O08779	O08779 rattus norv
36	97.5	5.3	1480	5 Q9V7F8	Q9v7f8 drosophila
37	97.5	5.3	1504	5 Q9XYV4	Q9xyv4 drosophila
38	97.5	5.3	1504	5 Q9V7F9	Q9v7f9 drosophila
39	96	5.2	364	11 O70509	O70509 rattus norv
40	96	5.2	742	4 Q9UJ36	Q9uj36 homo sapien
41	95.5	5.1	364	6 O97569	O97569 ceratotheri
42	95.5	5.1	721	13 Q91902	Q91902 xenopus lae
43	95	5.1	261	10 O24530	O24530 vitis vinif
44	94.5	5.1	1328	13 P79754	P79754 fugu rubrip
45	94	5.1	637	10 Q9SF50	Q9sf50 arabidopsis

ALIGNMENTS

RESULT	1
Q9RY3	PRELIMINARY;
ID	Q9RY3
AC	Q9RY3; PRELIMINARY; PRT; 897 AA.
DT	01-OCT-2000 (Tremblrel. 15, Created)
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE	CD44-LIKE PRECURSOR FELL.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Tao Q., Zhang W., Cao X.;
RT	"Molecular cloning and characterization of human FELL sharing homology with CD44.";
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF160476; AAF82398.1; -.
SQ	SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match	94.6%;	Score 1757;	DB 4;	Length 897;
Best Local Similarity	93.8%;	Pred. No. 8.6e-162;		
Matches 331;	Conservative	3;	Mismatches 19;	Indels 0; Gaps 0;
QY	1	MTGPGKHCKECKSHYVGDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL	60	
Db	489	MTGPGKHCKECKSHYVGDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL	548	
QY	61	RSPFGQYKLTFFDKARACANEAATMATYNQLSYOKAKYHLCSSAGWLETGRVAYPTAFAS	120	
Db	549	RSPFGQYKLTFFDKARACANEAATMATYNQLSYOKAKYHLCSSAGWLETGRVAYPTAFAS	608	
QY	121	QNCGSGVGVGVDPKPNKSEMDVFCYRMKDVNCTKXGVYVGDGFSYSGNLQVLMSEFP	180	
Db	609	QNCGSGVGVGVDPKPNKSEMDVFCYRMKDVNCTKXGVYVGDGFSYSGNLQVLMSEFP	668	
QY	181	SLTNFLTEVLAYSNSARGRAFLHLEHTDLSIRGTLFXPONSGLGENETLSGRDIEHHLAN	240	
Db	669	SLTNFLTEVLAYSNSARGRAFLHLEHTDLSIRGTLFXPONSGLGENETLSGRDIEHHLAN	728	

```
QY 241 VSMFFYNDLVNGTTLOTRGLSGKLLITDRQDPLHPTETRCVGDGRDITLWDICASNGITHVI 300
Db 729 VSMFFYNDLVNGTTLOTRVSGSKLLITASQDPLQPTETRFVDGRAILQWDFASNGIHHVI 788
QY 301 SRXLKAPPAPVTLXHTLXGIFXXIILVTGAVALAAYSFRINRRTIGFXHF 353
Db 789 SRPLKAPPAPVTLTHTGLGAGIFFAILLVTGAVALAAYSFRINRRTIGFQHF 841

RESULT 2
Q9UF98 PRELIMINARY; PRT; 1069 AA.
AC OSUF98
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).
GN DKF2P434B0321.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL13021; CAB61358.1; -.
DR HSSP; P98066; ITSG.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000782; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF00008; EGF; 8.
DR PRINTS; PF00193; Xlink; 1.
DR PRINTS; PR00111; EGFLAMININ.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1069 AA; 115666 MW; 04B0960932164D63 CRC64;

Query Match 79.1%; Score 1468; DB 4; Length 1069;
Best Local Similarity 80.5%; Pred. No. 1.2e-133;
Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHL 60
Db 697 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHL 756
QY 61 RSPGLQYKLTDFDKAREACANEATMATYNQLSYXQAKYHLCAGWLETGRVAYPTAFAS 120
Db 757 RSPGLQYKLTDFDKAREACANEATMATYNQLSYXQAKYHLCAGWLETGRVAYPTAFAS 816
QY 121 QNCGSGVGVIVDYGPRNPKSEMWDVFCYRMKDVNCTKXKVGVDGFS-YSGNLLQVLMSP 180
Db 817 QNCGSGVGVIVDYGPRNPKSEMWDVFCYRMK-----SAGLEFQQLSSRP 860
QY 181 SLTNFLTEVLAYSNSARGRAFLEHLTDLSTIRGTLFXPQNSGLGENETLSGRDIEHHLA 240
Db 861 CIS-----RTPDDLSTIRGTLFXPQNSGLGENETLSGRDIEHHLA 900
QY 241 VSMFFYNDLVNGTTLOTRGLSGKLLITDRQDPLHPTETRCVGDGRDITLWDICASNGITHVI 300
Db 901 VSMFFYNDLVNGTTLOTRGLSGKLLITASQDPLQPTETRFVDGRAILQWDFASNGIHHVI 960
QY 301 SRXLKAPPAPVTLXHTLXGIFXXIILVTGAVALAAYSFRINRRTIGFXHF 353
Db 961 SRPLKAPPAPVTLTHTGLGAGIFFAILLVTGAVALAAYSFRINRRTIGFQHF 1013
```

```
RESULT 3
Q93072 PRELIMINARY; PRT; 2212 AA.
AC Q93072;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MYELOBLAST KIAA0246 PROTEIN (FRAGMENT).
GN KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RL analysis of cDNA clones from cell line KG-1 and brain.";
DNA Res. 3:321-329(1996).
DR EMBL; D87433; BAA13377.1; -.
DR HSSP; P98066; ITSG.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000782; -.
DR INTERPRO; IPR001128; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF00008; EGF; 13.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 13.
DR PROSITE; PS01241; LINK; UNKNOWN_1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 2.
KW Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 2212 AA; 237451 MW; 4A95460504129134 CRC64;

Query Match 35.6%; Score 662; DB 4; Length 2212;
Best Local Similarity 40.2%; Pred. No. 4.9e-55;
Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGPGRKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDITVGVFHL 60
Db 1796 TGLNTRRCECHAGYVGDGLQCLESEPPVDRCLGQPPPCHSNDAMCTDLHFQEKRAGVFHL 1855
QY 61 RSPGLQYKLTDFDKAREACANEATMATYNQLSYXQAKYHLCAGWLETGRVAYPTAFAS 120
Db 1856 QATSPGYGLNFSEAEACAEAGAVLASPPQLSAAQQLGFLCLMGWLANGSTAHVPVFPV 1915
QY 121 QNCGSGVGVIVDYGPRNPKSEMWDVFCYRMKDVNCTKXKVGVDGFS-YSGNLLQVLMSP 179
Db 1916 ADCGNRGVIVSLGARKNLSEWDAYCFRQDVACRCRNGFVGDICTCNKGLDLVAAT 1975
QY 180 PSLTNFLTEVLAYSNSARGRAFLEHLTDLSTIRGTLFXPQNSGLGENETLSGRDIEHHLA 239
Db 1976 ANFSTFYGMLLGYANATQRLDLDLDELTYKTFLFVFNVEGVNDMTLSGPDLEHAS 2035
QY 240 NVSMFFYNDLVNGTTLOTRGLSGKLLITDR-----QDPLHPTETRCVGDGRDITLWDICASNGI 296
Db 2036 NATLISAN-ASQKLLPAHSLGLSLIISDAGPDNNSWAPVCTVVSRIIIVDLMFNGI 2094
QY 297 THVISRLKAPPAPVTLXHTLXGIFXXIILVTGAV--ALAAYSYFRINRRTIGF 350
Db 2095 IHALASPLAPPQPAVLAPPEPPVAAGVAVLAAGALLGLVAGALYLRRGKPMGF 2150
```

RX	MEDLINE=98087423; PubMed=9427551;
RA	Fulop C., Kamath R.V., Li Y., Otto J.M., Salustri A., Olsen B.R.,
RA	Giant T.T., Hascall V.C.;
RT	*Coding sequence, exon-intron structure and chromosomal localization
RT	of murine TNF-stimulated gene 6 that is specifically expressed by
RT	expanding cumulus cell-oocyte complexes.*;
RL	Gene 202:95-102(1997).
DR	EMBL; U83903; AAC53527.1; -.
DR	HSSP; P98066; ITSG.
DR	MGD; MGI:1195266; Tnfip6.
DR	INTERPRO; IPR000538; -.
DR	INTERPRO; IPR000859; -.
DR	PFAM; PF00193; Xlink; 1.
DR	PFAM; PF00431; CUB; 1.
DR	PROSITE; PS01180; CUB; 1.
DR	PROSITE; PS01241; LINK; 1.
DR	PRODOM; PD00918; -; 1.
SQ	SEQUENCE 275 AA; 30924 MW; 1CD247228260B8F9 CRC64;
Query Match 14.2%; Score 264; DB 11; Length 275;	
Best Local Similarity 43.3%; Pred. No. 1.2e-17;	
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps	
Qy	52 DTTGVFHLRSLGQYKTFDKARACANEATMATYNQLSYQKAKYHLCISAGWLETGR 111 : : :: : : :: : : : : : Db 32 EQAAGVYHREAGRKYKLTYAEAKAVCFEGGRLATYKQLEARKIGFHVCAGHWAKGR 91
Qy	112 VAYPTAFASONGCGVGIVDYGPRPNKSEMMDVFICYRMKDVCN 155 Db 92 VGYPVIRPGPCNGFGTKGIIDYGLRNRSERWDACYNPHAKEC 135
RESULT 6	
P79787	
ID	P79787 PRELIMINARY; PRT; 2109 AA.
AC	
DT	01-MAY-1997 (TrEMBLrel. 03, Created)
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
OS	CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN.
DS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN;
EX	MEDLINE=96262324; PubMed=8965652;
FX	Li H., Domowicz M., Hennig A., Schwartz N.B.;
RT	"S103L reactive chondroitin sulfate proteoglycan (aggrecan) mRNA
RT	expressed in developing chick brain and cartilage is encoded by a
RT	single gene.";
RL	Brain Res. Mol. Brain Res. 36:309-321(1996).
DR	EMBL; U78555; AAC60053.1; -.
DR	HSSP; P08709; IBF9.
DR	INTERPRO; IPR000152; -.
DR	INTERPRO; IPR000436; -.
DR	INTERPRO; IPR000538; -.
DR	INTERPRO; IPR000561; -.
DR	INTERPRO; IPR000742; -.
DR	INTERPRO; IPR001304; -.
DR	INTERPRO; IPR001881; -.
DR	INTERPRO; IPR003006; -.
DR	PFAM; PF00008; EGF; 1.
DR	PFAM; PF00047; ig; 1.
DR	PFAM; PF00059; lectin_c; 1.
DR	PFAM; PF00084; sushi; 1.
DR	PFAM; PF00193; Xlink; 4.
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR	PROSITE; PS00615; C_Type_I_FcRn_1; 1.


```

QY 55 VGVFHLRSLGQYKLTFFDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAY 114
DB 159 GVVFHRSARDRYALTFAEQAACRLSSAIIAAPHRLQAFAEDGPDNCDAGWLSDRTVRY 218
QY 115 PTFASQNC---GSGVVGVDYDGPKNKSEMDVFCYRMKDVNCTXXKGVYVG 163
DB 219 PITQSRPGCYGRDSSLPGVRSYG-RRNPQELDYVCFARE---LGGEVYVG 266

RESULT 9
ID O88564 PRELIMINARY; PRT; 655 AA.
AC O88564;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE VERISCAN V3 ISOFORM PRECURSOR.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=WISTAR KYOTO;
RA Lemire J.M., Braun K.R., Maurel P., Margolis R.U., Schwartz S.M.,
RA Wight T.N.;
RT "Verisican isoforms in vascular smooth muscle cells.";
RL EMBL; AF072892; AAC26116.1; -;
DR HSSP; P01132; 1EGF.
DR INTERPRO; IPR000152; -;
DR INTERPRO; IPR000436; -;
DR INTERPRO; IPR000538; -;
DR INTERPRO; IPR000361; -;
DR INTERPRO; IPR000742; -;
DR INTERPRO; IPR001304; -;
DR INTERPRO; IPR001438; -;
DR INTERPRO; IPR001881; -;
DR INTERPRO; IPR003006; -;
DR PFAM; PF00008; EGF_2; -;
DR PFAM; PF00047; Ig; 1; -;
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PRODOM; PD000918; -; 2.
KW Signal; Glycoprotein; EGF-like domain.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 655 VERISCAN V3 ISOFORM.
SQ SEQUENCE 655 AA; 74474 MW; 602F2F37E4F1ECC4 CRC64;

```

```

Query Match 8.7%; Score 161.5; DB 11; Length 655;
Best Local Similarity 26.1%; Pred. No. 3.5e-07;
Matches 73; Conservative 30; Mismatches 98; Indels 79; Gaps 13;

QY 57 VFHLRSLGQYKLTFFDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAYPT 116
DB 151 VFHYRAATSYTLNFESAQAACLDIGAVIATPELFAAAYEDGFEQCDAGWLSQTVRYPI 210
QY 117 AFASQNCGS---GVGVGVDYDGPKNKSEMDVFCYRMKDVNCTXXKGVY-GDGFYSGNL 172
DB 211 RAPREGCYGDMGKGVRYGFR-SQETIYDVICY-----VDHLDGDFVH----- 254
QY 173 LQVLMSPSLTNFLTEVLAYSNSARGRAFLE-----HLTDLISRG--TL 215

```

```

DB 255 ----ITAPSKFTFEAEAEACANRDLARLATVGEHHAARNRNGEQDQYGLSDASVRRHPVT 310
QY 216 FXPNQSG--LG-----ENETL-----SGRDIEHLANVSFFYNDLVNGTTLQTRLGSK 262
DB 311 ARAQCQGGLLGVRTLRYFENQTCFPLPDSRFDAYCFKRPDLCKTNPCNLNGGT----- 362
QY 263 LLITRDQDPLHPTET-----RCVDG-----RDTLEWDICASN 294
DB 363 -----CYPTETSYVCTCAPGYSGDQCELDLDFDECHSN 393

RESULT 10
Q9W6E1 PRELIMINARY; PRT; 1290 AA.
ID Q9W6E1;
AC Q9W6E1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE NEUROCAN CORE PROTEIN PRECURSOR.
DE NEUROCAN CORE PROTEIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RA Li H., Leung T., Balsamo J., Hoffman S., Lillien J.;
RT "cDNA cloning of chicken neurocan and its role in regulating N-
cadherin function in embryonic chicken retina.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF116856; AAD24546.2; -;
DR HSSP; P08709; 1BF9.
DR INTERPRO; IPR000152; -;
DR INTERPRO; IPR000436; -;
DR INTERPRO; IPR000538; -;
DR INTERPRO; IPR000561; -;
DR INTERPRO; IPR000742; -;
DR INTERPRO; IPR001304; -;
DR INTERPRO; IPR001438; -;
DR INTERPRO; IPR001881; -;
DR PFAM; PF00008; EGF_2; -;
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1290 AA; 138876 MW; 182BD86D0E40BE78 CRC64;

```

```

Query Match 8.7%; Score 161.5; DB 13; Length 1290;
Best Local Similarity 29.1%; Pred. No. 8.6e-07;
Matches 43; Conservative 21; Mismatches 51; Indels 33; Gaps 5;

QY 5 GKHKCECKSHYVGLNCEPEQLPIDRCLQDNQCHADKCVDLHFQDTYVG-VFHLRSP 63
DB 136 GLYRCE-----VVAGIDDDSDLLPLE-----VMGVVFHYRPA 167
QY 64 LGQYKLTFFDKAREACANEATMATYNQLSYXOKAKYHLCSAGWLETGRVAYPTAFASQNC 123
DB 168 GARYALTFAAARRACDNSAVIASPOHLQAFAEDGVDNCDAGWLDQSVRYPTILSRPGC 227
QY 124 ---GSGVGVGVDYDGPKNKSEMDVFCY 148
DB 228 YGDRNSLPGVRSYGQR-EPGELYDVICY 254

```

```
RESULT 11
O77610 ID O77610 PRELIMINARY; PRT; 2394 AA.
AC O77610;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE VERSICAN V1 SPLICE-VARIANT PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9828320; Pubmed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain."
RL J. Biol. Chem. 273:15758-15764 (1998).
DR EMBL; AF060457; AAC24359.1; -.
DR HSSP; P01132; LEFG.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000436; -.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001304; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00008; EGF; 2.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00084; sushi; 1.
DR PFAM; PF00193; Xlink; 2.
DR PRINTS; PRO0010; EGFLOOD.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR PRODOM; PD000918; -.
KW Signal; Glycoprotein; EGF-like domain.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2394 VERSICAN V1 SPLICE-VARIANT.
FT SEQUENCE 2394 AA; 261886 MW; B82A3E10FC5BD990 CRC64;

Query Match 8.6%; Score 160.5; DB 6; Length 2394;
Best Local Similarity 25.5%; Pred. No. 2.4e-06;
Matches 83; Conservative 31; Mismatches 113; Indels 99; Gaps 17;

Qy 51 QDT---TVG---VHLRSLPGQYKLTDKAREACANEATMYNQLSYQKAKYHLCSAG 105
Db 141 QDVTSLTVEGVFHYRAATSRVTLNFEAMQKACVDIGAVIATPEQLHAAAYEDGFEQCDAG 200
Qy 106 WLETGRVAYPTAFASONGS---GVGVGVDYDGPKNKSEMDVFCY----- 148
Db 201 WLSDQTVRYPIRVPREGCYGDMGKGVRTYGFPA-PHETIDYVCYVDHLDGDFVHTAP 259
Qy 149 -----RMKDVNCTXKGVYGD-----GF-----S 167
Db 260 NKFTFEAGEECKTQD----ARLATVGELOAAWRNGFDRCDYGWLLDASVRHPTVARAQ 315
Qy 168 YSNNLLOV--LMFSPSLTNPLTEVLAYSNSARGAFLEHLTDLSTIRGLTFXPQNSGLGE 225
Db 316 CGGGLLGVRTLYRFENOTGFPT-----PDSRFDAYCFKRRMSDFSVSG---HPIDSESKE 367

Qy 226 NETLSGR-DIEHHLANVSMFFYNDLVNGTTLQTRLSGSKLLITRDQDPLHPTETRCVGDGRD 284
Db 368 DEPCSEETDPEH-----DLI-AETPELLGLMLHSEDEED-----EECANATD 409
Qy 285 -TLEWDICASNGITHVISRLKAPPA 309
Db 410 VTTTPSVQYING-KHVVTTPKDPPEA 434

RESULT 12
O02817 ID O02817 PRELIMINARY; PRT; 192 AA.
AC O02817;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE LARGE AGGREGATING CARTILAGE PROTEOGLYCAN CORE PROTEIN (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RA Nishimura M., Noshiro M., Kawamoto T., Nakamasu K., Hamada T.,
RA Kato Y.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB004812; BAA20524.1; -.
DR HSSP; P98066; 1TSG.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS01241; LINK; 1.
FT NON_TER 1 192
FT SEQUENCE 192 AA; 21606 MW; 44EA35FA92CEB8CC CRC64;

Query Match 8.2%; Score 153; DB 6; Length 192;
Best Local Similarity 35.8%; Pred. No. 4.5e-07;
Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps 2;

Qy 57 VHLRSLPGQYKLTDKAREACANEATMYNQLSYQKAKYHLCSAGHLETGRVAYPT 116
Db 97 VFHYRAISTRYTLDFDRAQRAQLNSAIATPEQLQAAAYEDGFHQCDAGWLADQTVRYPI 156
Qy 117 AFASONGS---GVGVGVDYDGPKNKSEMDVFCY 148
Db 157 HTPREGCYGDKDEFPVGVRTYGIIR-DTNETYDVICYF 190

RESULT 13
O77612 ID O77612 PRELIMINARY; PRT; 656 AA.
AC O77612;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE VERSICAN V3 SPLICE-VARIANT PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9828320; Pubmed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
```


Query Match	8.2%	Score 151.5;	DB 6;	Length 1643;
Best Local Similarity	36.8%;	Pred. No. 1.1e-05;		
Matches	39;	Conservative	13; Mismatches	45; Indels 9; Gaps

QY	51	QDTV--TVG-VFHLRSLPGQYKLTEDKAREACAAATMATYNOLSYXOKAKYHLCGAS 105
Db	141	QDTVSITVEGVVFHYRAATSRYTLNFEMAQKACVDIGAVIATPEQLHAYIEDGEFQCDAQ 200
		: : : : : : : : : : : : : :
QY	106	WLETGRVAYPTAFASQNGS---GVVGIVDYGPRPNKSEMMWDVFCY 148
Db	201	WLSQDTVRYPPIRPREGCYGDMMGKEGRTYGFPA-PHETYDVICY 245
		: : : : : : : : : : : : :

RESULT	15
O77609	PRELIMINARY; PRT; 3381 AA.
ID	O77609
AC	01-NOV-1998 (TRENBLrel. 08, Created)
DT	01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DE	01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE	VERSCAN V0 SPICE-VARIANT PRECURSOR.
OS	Bos taurus (Bovine).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX	Bovidae; Bovinae; Bos.
OX	NCBI_TaxId=9913;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98288320; PubMed=9624174;
RA	Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA	Zimmermann D.R.;
RT	"Versican V2 is a major extracellular matrix component of the mature
RT	bovine brain.";
J.	J. Biol. Chem. 273:15758-15764(1998).
DR	EMBL: AF060456: AAC24358.1;

```

DR HSP; P01132; LEGF.
DR INTERPRO: IPR000152; -
DR INTERPRO: IPR000436; -
DR INTERPRO: IPR000538; -
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR000742; -
DR INTERPRO: IPR001304; -
DR INTERPRO: IPR001438; -
DR INTERPRO: IPR001881; -
DR INTERPRO: IPR003006; -
DR PFAM: PF000008; EGF; 2.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00059; lectin_c; 1.
DR PFAM: PF00084; sushi; 1.
DR PFAM: PF00193; xlnk; 2.
DR PRINTS; PR00010; EGFBLD.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR PRODOM; PD000918; -; 2.
KW Signal; Glycoprotein; EGF-like domain.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 VERSICAN V0 SPLICE-VARIANT.
SQ SEQUENCE 3381 AA; 369987 MW; F09716FA7778D459 CRC64;

Query Match 8.2%; Score 151.5; DB 6; Length 3381;
Best Local Similarity 36.8%; Pred. No. 2.9e-05;
Matches 39; Conservative 13; Mismatches 45; Indels 9; Gaps 4;

QY 51 QDT---TVG--VFHLRPLGQYKLTDFKAREACANEAAATMATYNQLSYXOKAKYHLCASG 105
Db ||| ||| | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 141 QDTVSLTVEGVFHYRAATSRVTLNFEAMQKACVDIGAVIATPEQLHAAAYEDGFEQCDA 200

QY 106 WLETGRVAYTATASQNCGS---GVVGIVDYGPRPNKSEMWDFCY 148
Db ||| ||| | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 201 WLSQDTVRYPIRVPRGCGYDMMGKGVRTYGFRA-PHETYDVICY 245

```

Search completed: April 4, 2001, 13:06:28
Job time: 147 sec

Handwritten:
New found
found

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:21 ; Search time 11.97 Seconds
(without alignments)
952.365 Million cell updates/sec

Title: US-09-466-778-11
Perfect score: 1857
Sequence: 1 MTGPGKHCKECKSHYVGDGL.....ALAAYSYFRINKTIGFXHF 353

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	264	14.2	276	1 TSG6_RABIT	P98065 oryctolagus
2	262	14.1	277	1 TSG6_HUMAN	P98066 homo sapien
3	175	9.4	2109	1 PGCA_CHICK	P07898 gallus gall
4	171	9.2	2124	1 PGCA_RAT	P07897 rattus norv
5	170	9.2	2364	1 PGCA_BOVIN	P13608 bos taurus
6	169.5	9.1	354	1 PLK_RAT	P03994 rattus norv
7	169.5	9.1	356	1 PLK_MOUSE	Q9qub5 mus musculu
8	166.5	9.0	354	1 PLK_HORSE	Q28381 equus cabal
9	165.5	8.9	354	1 PLK_BOVIN	P52522 bos taurus
10	165.5	8.9	355	1 PLK_CHICK	P07354 gallus gall
11	165	8.9	2132	1 PGCA_MOUSE	Q61282 mus musculu
12	163.5	8.8	354	1 PLK_HUMAN	P10915 homo sapien
13	161.5	8.7	354	1 PLK_PIG	P10859 sus scrofa
14	161.5	8.7	2415	1 PGCA_HUMAN	P16112 homo sapien
15	158.5	8.5	1268	1 PCGN_MOUSE	P55066 mus musculu
16	157.5	8.5	883	1 PGCB_RAT	P55068 rattus norv
17	157.5	8.5	1257	1 PCGN_RAT	P58067 rattus norv
18	155.5	8.4	394	1 PGCA_RABIT	P28670 oryctolagus
19	155.5	8.4	912	1 PGCB_BOVIN	Q28062 bos taurus
20	154.5	8.3	862	1 PCGV_MACNE	Q28858 macaca neme
21	152.5	8.2	3358	1 PCGV_MOUSE	Q62059 mus musculu
22	152.5	8.2	3562	1 PCGV_CHICK	Q29053 gallus gall
23	150.5	8.1	883	1 PGCB_MOUSE	Q61361 mus musculu
24	150.5	8.1	3396	1 PCGV_HUMAN	P13611 homo sapien
25	147	7.9	417	1 PGCB_FELCA	P41725 felis silve
26	111.5	6.0	362	1 CD44_CRIGR	P20944 cricetus
27	108.5	5.8	431	1 CD44_MESAU	Q60522 m cd44 anti
28	106	5.7	359	1 CD44_HORSE	Q05078 equus cabal
29	105	5.7	810	1 NELL_HUMAN	Q28832 homo sapien
30	103.5	5.6	655	1 CD44_MOUSE	P15379 mus musculu
31	100	5.4	351	1 CD44_CANFA	Q28284 canis fami
32	99	5.3	2907	1 FBN2_MOUSE	Q61555 mus musculu
33	98	5.3	742	1 CD44_HUMAN	P16070 h cd44 anti

34	97.5	5.3	503	1	CD44_RAT	P26051 rattus norv
35	97.5	5.3	1480	1	SLIT_DROME	P24014 drosophila
36	97	5.2	2911	1	FBN2_HUMAN	P35556 homo sapien
37	96	5.2	713	1	TSA4_GIALA	P21849 giardia lam
38	95.5	5.1	366	1	CD44_BOVIN	Q29423 bos taurus
39	95.5	5.1	683	1	BGH3_PIG	O11780 sus scrofa
40	94.5	5.1	722	1	DL11_MOUSE	Q61483 mus musculu
41	94.5	5.1	1375	1	NID2_HUMAN	Q14112 homo sapien
42	93.5	5.0	723	1	DL11_HUMAN	Q00548 homo sapien
43	93	5.0	1245	1	NIDO_MOUSE	P10493 mus musculu
44	92.5	5.0	362	1	CD44_PAPHA	P14745 papio hamad
45	92.5	5.0	1403	1	NID2_MOUSE	O88322 mus musculu

ALIGNMENTS

RESULT	1
TSG6_RABIT	
ID	TSG6_RABIT
AC	P98065; STANDARD; PRT; 276 AA.
DT	01-FEB-1996 (Rel. 33, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN TSG-6 PRECURSOR (HYALURONATE- BINDING PROTEIN PS4)
DE	TNFAIP6 OR TSG6 OR PS4.
GN	Oryctolagus cuniculus (Rabbit).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NEW ZEALAND WHITE;
RX	MEDLINE=93252803; PubMed=8098034;
RA	Feng P., Liaw G.;
RT	"Identification of a novel serum and growth factor-inducible gene in vascular smooth muscle cells.";
RL	J. Biol. Chem. 268:9387-9392(1993).
RN	[2]
RP	ERRATUM.
RX	MEDLINE=94012707; PubMed=8407990;
RA	Feng P., Liaw G.;
RL	J. Biol. Chem. 269:21453-21453(1993).
CC	-!- FUNCTION: POSSIBLY INVOLVED IN CELL-CELL AND CELL-MATRIX INTERACTIONS DURING INFLAMMATION AND TUMORGENESIS.
CC	-!- TISSUE SPECIFICITY: VASCULAR SMOOTH MUSCLE CELLS.
CC	-!- DEVELOPMENTAL STAGE: FETAL.
CC	-!- INDUCTION: BY SERUM AND GROWTH FACTOR.
CC	-!- SIMILARITY: CONTAINS 1 LINK DOMAIN.
CC	-!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M86381; AAA03342.1; -
DR	HSSP; P98066; 1TSG.
DR	INTERPRO; IPR000538; -
DR	INTERPRO; IPR000859; -
DR	PFAM; PF00431; CUB; 1.
DR	PFAM; PF00193; Xlink; 1.
DR	PROSITE; PS01180; CUB; 1.
DR	PROSITE; PS01241; LINK; 1.
DR	Cell adhesion; Signal; Glycoprotein.
FT	POTENTIAL.
FT	CHAIN 1 19
FT	20 276
FT	TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN
FT	TSG-6.
FT	DOMAIN 53 128
FT	LINK.
FT	DOMAIN 135 247
FT	CUB.

```
FT DISULFID 58 127 BY SIMILARITY.
FT DISULFID 82 103 BY SIMILARITY.
FT DISULFID 135 161 BY SIMILARITY.
FT DISULFID 188 210 BY SIMILARITY.
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 276 AA; 31081 MW; 3BDC5D9A24B2F75A CRC64;

Query Match 14.2%; Score 264; DB 1; Length 276;
Best Local Similarity 43.3%; Pred. No. 2.5e-16;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTVGVFHLRSPGLQYKLTDFKAREACANEATMATYNQLSYQKAKYHLCAGWLETGR 111
Db 32 EQAAGVYHREARSGKYKLTAEAKAVCEFEGLRLATYKQLEAAKRGIFHVCAAGNMAKGR 91

Qy 112 VAYPTAFASONGSGVGVGDYGRPNKSEMDVFCYRMKDVC 155
Db 92 VGYPIVPGSGFGKGTGIIDYGIRLNRSEMDAYCYNPHAKEC 135

RESULT 2
TSQ6_HUMAN STANDARD; PRT; 277 AA.
AC P98066;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN TSG-6 PRECURSOR (HYALURONATE-
DE BINDING PROTEIN).
GN TNFAIP6 OR TSG6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP "A novel secretory tumor necrosis factor-inducible protein (TSG-6) is
RP a member of the family of hyaluronate binding proteins, closely
RP related to the adhesion receptor CD44."
RL J. Cell Biol. 116:545-557(1992).
RN [2]
RP STRUCTURE BY NMR OF 36-133.
RX MEDLINE=96390850; PubMed=8797823;
RA Kohda D., Morton C.J., Parker A.A., Hatanaka H., Inagaki F.M.,
RA Campbell I.D., Day A.J.;
RT "Solution structure of the link module: a hyaluronan-binding domain
RT involved in extracellular matrix stability and cell migration."
RL Cell 86:767-775(1996).
CC -!- FUNCTION: POSSIBLY INVOLVED IN CELL-CELL AND CELL-MATRIX
CC INTERACTIONS DURING INFLAMMATION AND TUMORGENESIS.
CC -!- INDUCTION: BY TNF.
CC -!- SIMILARITY: CONTAINS 1 LINK DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01241; LINK; 1.
KW Cell adhesion; Signal; Glycoprotein; 3D-structure.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 277 TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN
FT TSG-6.
FT DOMAIN 53 128 LINK.
FT DOMAIN 135 247 CUB.
FT DISULFID 58 127
FT DISULFID 82 103
FT DISULFID 135 161
FT DISULFID 188 210 BY SIMILARITY.
FT CARBOHYD 118 118 BY SIMILARITY.
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 277 AA; 31231 MW; 4DB3AE4AC52B880 CRC64;

Query Match 14.1%; Score 262; DB 1; Length 277;
Best Local Similarity 43.3%; Pred. No. 3.7e-16;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Qy 52 DTTVGVFHLRSPGLQYKLTDFKAREACANEATMATYNQLSYQKAKYHLCAGWLETGR 111
Db 32 ERAAGVYHREARSGKYKLTAEAKAVCEFEGLRLATYKQLEAAKRGIFHVCAAGNMAKGR 91

Qy 112 VAYPTAFASONGSGVGVGDYGRPNKSEMDVFCYRMKDVC 155
Db 92 VGYPIVPGSGFGKGTGIIDYGIRLNRSEMDAYCYNPHAKEC 135

RESULT 3
PGCA_CHICK STANDARD; PRT; 2109 AA.
ID PGCA_CHICK
AC P07898; Q90991; Q90820; Q91047; Q90810;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
GN AGCI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN; TISSUE=EMBRYO;
RX MEDLINE=94043149; PubMed=8226878;
RA Li H., Schwartz N.B., Vertel B.M.;
RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core
RT protein and identification of a stop codon in the aggrecan gene
RT associated with the chondrodystrophy, nanomelia."
RL J. Biol. Chem. 268:23504-23511(1993).
RN [2]
RP SEQUENCE OF 1042-1559 FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=90307744; PubMed=1694853;
RA Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II.
RT Nucleotide sequence of cDNA clone and localization of the S103L
RT epitope."
RL J. Biol. Chem. 265:12088-12097(1990).
RN [3]
RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
RC TISSUE=CARTILAGE;
RX MEDLINE=93111968; PubMed=1339285;
RA Chandrasekaran L., Tanzer M.L.;
RT "Molecular cloning of chicken aggrecan. Structural analyses."
RL Biochem. J. 288:903-910(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=94107258; PubMed=8280087;
RA Chandrasekaran L., Tanzer M.L.;
```

RL Biochem. J. 296:885-887(1993).
RN [5]
RP SEQUENCE OF 1492-1610 FROM N.A.
RC STRAIN-WHITE LEGRON; TISSUE=CHONDROCYTES;
RX MEDLINE=95128519; PubMed=7827752;
RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
RT "Molecular basis of nanomelia, a heritable chondrodystrophy of
RT chicken.";
RL Matrix Biol. 14:297-305(1994).
RN [6]
RP SEQUENCE OF 1894-2109 FROM N.A.
RX MEDLINE=89008500; PubMed=3170613;
RA Tanaka T., Har-El R., Tanzer M.L.;
RT "Partial structure of the gene for chicken cartilage proteoglycan
RT core protein.";
RL J. Biol. Chem. 263:15831-15835(1988).
RN [7]
RP SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
RX MEDLINE=86259736; PubMed=3460082;
RA Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;
RT "Cloning and sequence analysis of a partial cDNA for chicken
RT cartilage proteoglycan core protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -!- DISEASE: DEFECTS IN AGC1 IS THE CAUSE OF NANOMELIA, A LETHAL
CC CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT
CC (CHONDRODYSPLASIA) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS.
CC AGGREGAN IS TRUNCATED AT IS C-TERMINAL IN THE CS-2 BINDING DOMAIN
CC AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L21913; AAB19128.1; -;
DR EMBL; M38187; AAA48731.1; -;
DR EMBL; M88101; -; NOT_ANNOTATED_CDS.
DR EMBL; S74657; AAC60751.1; -;
DR EMBL; S74656; AAC60751.1; JOINED.
DR EMBL; J04028; AAA48719.1; -;
DR EMBL; M13993; AAA48720.1; -;
DR PIR; A25442; A25442.
DR HSPF; P00740; IIXA.
DR INTERPRO; IPR000152; -;
DR INTERPRO; IPR000436; -;
DR INTERPRO; IPR000538; -;
DR INTERPRO; IPR000561; -;
DR INTERPRO; IPR001304; -;

DR INTERPRO; IPR001881; -;
DR PFAM; PF00008; EGF; 1.
DR PFAM; PF00193; Xlink; 4.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 4.
DR PROSITE; PS0615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16
FT CHAIN 17 2109
FT DOMAIN 44 136
FT DOMAIN 166 243
FT DOMAIN 284 346
FT DOMAIN 537 614
FT DOMAIN 635 716
FT REPEAT 1363 1742
FT REPEAT 1855 1892
FT DOMAIN 1901 2019
FT REPEAT 2023 2081
FT DOMAIN 48 137
FT DOMAIN 148 243
FT DOMAIN 249 346
FT DOMAIN 519 613
FT DOMAIN 620 715
FT DOMAIN 718 803
FT DOMAIN 805 1264
FT DOMAIN 1265 1742
FT DOMAIN 1893 2109
FT DISULFID 51 129
FT DISULFID 171 242
FT DISULFID 195 246
FT DISULFID 289 314
FT DISULFID 293 314
FT DISULFID 542 613
FT DISULFID 566 587
FT DISULFID 640 715
FT DISULFID 664 685
FT DISULFID 1859 1870
FT DISULFID 1864 1879
FT DISULFID 1881 1890
FT DISULFID 1897 1908
FT DISULFID 1925 2017
FT DISULFID 1933 2009
FT DISULFID 2024 2067
FT DISULFID 2053 2080
FT CARBOHYD 76 76
FT CARBOHYD 122 122
FT CARBOHYD 330 330
FT CARBOHYD 388 388
FT CARBOHYD 439 439
FT CARBOHYD 644 644
FT CARBOHYD 700 700
FT CARBOHYD 765 765
FT CARBOHYD 801 801
FT VARSPIC 1856 1892
FT CONFLICT 362 362
FT CONFLICT 601 601
FT CONFLICT 1000 1000
FT CONFLICT 1029 1029
FT CONFLICT 1042 1043
FT CONFLICT 1251 1251
FT CONFLICT 1587 1587
FT CONFLICT 1590 1590
FT CONFLICT 1594 1594
FT CONFLICT 1602 1610
FT CONFLICT 1603 1603
FT CONFLICT 1672 1672
FT CONFLICT 1796 1796
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
MISSING (IN ISOFORM 2).
E -> D (IN REF. 3).
G -> D (IN REF. 3).
P -> R (IN REF. 3).
A -> P (IN REF. 3).
VT -> PA (IN REF. 2).
E -> D (IN REF. 2 AND 3).
I -> T (IN REF. 5).
I -> V (IN REF. 5).
T -> S (IN REF. 5).
IETSTVREI -> VLRCGSVLR (IN REF. 5).
E -> A (IN REF. 3).
S -> G (IN REF. 3).
E -> G (IN REF. 3 AND 7).

```
FT CONFLICT 1988 1988 F -> S (IN REF. 6).
SQ SEQUENCE 2109 AA; 223492 MW; 7F824FD5B3A2ABDA CRC64;

Query Match 9.4%; Score 175; DB 1; Length 2109;
Best Local Similarity 23.4%; Pred. No. 2.4e-07;
Matches 75; Conservative 31; Mismatches 116; Indels 98; Gaps 12;

Qy 10 ECKSHYVGDGLNCPPEQLPDRCLQDNGQCHA-----DAKVDLHFQDTTVG 56
Db 194 QCDAGLWADQVRYPPIHLPRPCRGDDEFPGVRYGVRETFDYVYCAEQMGK--- 250
Qy 57 VFHLRSLQGYKLFDRKARACANEATMATYNLSYQXQAKYHLCAGWLETGRVAYPT 116
Db 251 VFYATSP---EKFTFQAFDKCHSLGARLATTGELYLAWGDMDCSAGWLADRSVYPI 307
Qy 117 AFASONGCGVVGII--VDYGR-----PNKSEMDVFCYRKMKNVNC-----YXKVG-YV 162
Db 308 SRARPNCGGNLVGVRTVYLPANQTVGPHPSRYDAICYSGDDEPEALVPGLTDEVGTTEL 367
Qy 163 GDGFSY-----SCNLQVLMSPPSLTNFL 186
Db 368 GSAFTIQTQVTEVELPLPRNVNTEEEARGSIATLEPMEITATLEYEAFVLPDL--FA 425
Qy 187 TEVLAYSNSARGRAFLHLTDL-----SIRGTLFXPQNSGLGE----- 225
Db 426 TSVIVETASPREENVNTEETGCIWAVPEVTSVSGTAF---TTGMAEVSSVEEAIAYTA 482
Qy 226 ---NETLSGRDIEHHLNVS 242
Db 483 TPGLESASPFTIEDHLVQVT 502

RESULT 4
PGCA_RAT STANDARD; PRT; 2124 AA.
AC P07897;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
GN AGC1 OR AGC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RN J. Biol. Chem. 262:17757-17767(1987).
RN REVISION TO 698.
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RA J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RN SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86087070; PubMed=3693370;
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Sasaki M., Yamada Y.;
RA MEDLINE=86250698; PubMed=2424893;
RT "Complete primary structure of the rat cartilage proteoglycan core
RT protein deduced from cDNA clones."
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RN REVISION TO 698.
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RA J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RN SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus
RT of the rat cartilage proteoglycan."
RL J. Biol. Chem. 261:8108-8111(1986).
CC CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC • SIMILARITY).
CC CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC
```

```
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: M13518; AAA41836.1; -.
CC DR EMBL: J03485; AAA21000.1; ALT_SEQ.
CC DR PIR: A23835; A23835.
CC DR PIR: A28452; A28452.
CC DR HSSP: P98066; LTSG.
CC DR INTERPRO: IPR000436; -.
CC DR INTERPRO: IPR000495; -.
CC DR INTERPRO: IPR000538; -.
CC DR INTERPRO: IPR001304; -.
CC DR PFAM: PF00193; Xlink; 4.
CC DR PFAM: PF00059; lectin_c; 1.
CC DR PFAM: PF00084; sush1; 1.
CC DR PROSITE: PS00290; IG_MHC; 1.
CC DR PROSITE: PS01241; LINK; 4.
CC DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
CC DR PROSITE: PS00411; C_TYPE_LECTIN_2; 1.
CC KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
CC Repeat; Immunoglobulin domain.
CC SIGNAL 1 19 POTENTIAL.
CC FT CHAIN 20 2124 AGGREGAN CORE PROTEIN.
CC FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
CC FT DOMAIN 170 247 LINK 1.
CC FT DOMAIN 268 349 LINK 2.
CC FT DOMAIN 504 581 LINK 3.
CC FT DOMAIN 602 683 LINK 4.
CC FT DOMAIN 1910 2036 C-TYPE LECTIN.
CC FT REPEAT 2040 2098 SUSHI.
CC FT DOMAIN 48 140 GI-A.
CC FT DOMAIN 152 247 GI-B.
CC FT DOMAIN 253 349 GI-B'.
CC FT DOMAIN 486 580 G2-B.
CC FT DOMAIN 587 682 G2-B'.
CC FT DOMAIN 685 798 KS.
CC FT DOMAIN 801 1226 CS-1.
CC FT DOMAIN 1227 1909 CS-2.
CC FT DOMAIN 1910 2124 G3.
CC FT DISULFID 51 133 BY SIMILARITY.
CC FT DISULFID 175 246 BY SIMILARITY.
CC FT DISULFID 199 220 BY SIMILARITY.
CC FT DISULFID 273 348 BY SIMILARITY.
CC FT DISULFID 297 318 BY SIMILARITY.
CC FT DISULFID 509 580 BY SIMILARITY.
CC FT DISULFID 533 554 BY SIMILARITY.
CC FT DISULFID 607 682 BY SIMILARITY.
CC FT DISULFID 631 652 BY SIMILARITY.
CC FT DISULFID 1914 1925 BY SIMILARITY.
CC FT DISULFID 1942 2034 BY SIMILARITY.
CC FT DISULFID 2010 2026 BY SIMILARITY.
CC FT DISULFID 2041 2084 BY SIMILARITY.
CC FT DISULFID 2070 2097 BY SIMILARITY.
CC FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
```

```
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1842 1842 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2124 AA; 221117 MW; E30BB61593A34B1 CRC64;

Query Match 9.28; Score 171; DB 1; Length 2124;
Best Local Similarity 28.88; Pred. No. 5.6e-07;
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;

OY 57 VPHLSPLGQYKLTDPKAREACANEATMATYNQLSYQKAKYHLCSAGWLETGRVAYPT 116
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 488 VFHYRPGSTRVSLTFEEAQACIRFGAASPAEQAAEAGYEQCDAGWLDQOTVRPI 547
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
OY 117 AFASQNC---GSGVVGIVDYGPRPNKSEMDVFCYRKMKNVCTKVGVDGFSYSGNLL 173
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 548 VSPRTPCVGDKDSFGVRYGVRPS-SETYDVYCYVDK-----LEGEVF----- 590
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
OY 174 QVLMFSPSLTNLTFLVAYSSSARGRAFLHLTDLSTGLTFXPQNSGL 223
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 591 -----FATQMEQFTFEAQAFCAQNAAT-LASTGQLYAAWSQGL 628
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 5
PGCA_BOVIN STANDARD; PRT: 2364 AA.
ID PGCA_BOVIN STANDARD; PRT: 2364 AA.
AC P13608; Q28159; P79117;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
GN AGC1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RA Hering T.M., Kollar J., Huynh T.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 563-1056 FROM N.A.
RX MEDLINE-89380219; PubMed-2528543;
RA Antonsson P., Heinegaard D., Oldberg A.;
RT "The keratan sulfate-enriched region of bovine cartilage proteoglycan
RT consists of a consecutively repeated hexapeptide motif.";
RL J. Biol. Chem. 264:16170-16173(1989).
RN [3]
RN SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RX MEDLINE-87270630; PubMed-3111460;
RA Oldberg A., Antonsson P., Heinegaard D.;
RT "The partial amino acid sequence of bovine cartilage proteoglycan,
RT deduced from a cDNA clone, contains numerous Ser-Gly sequences
RT arranged in homologous repeats.";
RL Biochem. J. 243:255-259(1987).
RN [4]
RN SEQUENCE OF 2114-2150 FROM N.A.
RC TISSUE-CARTILAGE;
RX MEDLINE-93352525; PubMed-8349621;
RA Fueleop E., Walcz E., Vallyon M., Glant T.T.;
RT "Expression of alternatively spliced epidermal growth factor-like
RT domains in aggrecans of different species. Evidence for a novel
RT module.";
RL J. Biol. Chem. 268:17377-17383(1993).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE-85027710; PubMed-6489519;
RA Perin J.P., Bonnet F., Jolles J., Jolles P.;
RT "Sequence data concerning the protein core of the cartilage
RT proteoglycan monomers. Characterization of a sequence allowing the
```

```
RT synthesis of an oligonucleotide probe.";
RL FEBS Lett. 176:37-42(1984).
RN [6]
RP PARTIAL SEQUENCE.
RX MEDLINE-87005253; PubMed-3530809;
RA Perin J.P., Bonnet F., Jolles P.;
RT "Structural relationship between link proteins and proteoglycan
RT monomers.";
RL FEBS Lett. 206:73-77(1986).
CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC -!- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN
CC ADULT AND FETAL BOVINE PROTEOGLYCAN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U76615; AAB38524.1; -.
DR EMBL; L07053; -. NOT_ANNOTATED_CDS.
DR PIR; A2752; A27752.
DR PIR; A29164; A29164.
DR PIR; B29164; B29164.
DR PIR; E29164; E29164.
DR PIR; G27751; G27751.
DR HSP; P00740; IIXA.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000436; -.
DR INTERPRO; IPR000538; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR001304; -.
DR PFAM; PF00008; EGF; 1.
DR PFAM; PF00193; Xlink; 4.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sush; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 4.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00411; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 2364 AGGREGAN CORE PROTEIN.
FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
```



```
Db 135 GRYKCE-----VIEGL-----EDDTAVVALDLQGV---VFPYFRL 167
QY 65 GOYKLTEDKAREACANEATMATYNOLSYQXKAKYHLCAGWLETGRVAYPTAFASQNG 124
Db 168 GRYNLNFHEAQACLDQDAVIASFQDLYDAWRGLDWCNAGWLSGDSGVQYPTTKPRECG 227
QY 125 --SGVVGVIVDYGPRPNKSEMMDVFCY 148
Db 228 GQNTVPGVRNYGFWDKSKSYDVFCF 253

RESULT 9
PLK_BOVIN
ID PLK_BOVIN STANDARD; PRT; 354 AA.
AC P55252;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).
GN CRTLL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RX MEDLINE=96043243; PubMed=7584851;
RA Hering T.M., Kollar J., Huynh T.D., Sandell L.J.;
RT "Bovine chondrocyte link protein cDNA sequence: interspecies
  conservation of primary structure and mRNA untranslated regions.";
RL Comp. Biochem. Physiol. 112B:197-203(1995).
CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
  HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U02292; AAC04311.1; -
CC HSP; P98066; ITS.
CC INTERPRO; IPR000538; -
CC INTERPRO; IPR003006; -
CC PFAM; PF00193; Xlink; 2.
CC PFAM; PF00047; ig; 1.
CC PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT PROPEP 1 15 BY SIMILARITY.
FT CHAIN 16 354 PROTEOGLYCAN LINK PROTEIN.
FT DOMAIN 54 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 350 LINK 2.
FT DISULFID 61 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 349 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
SQ SEQUENCE 354 AA; 40287 MW; 065D155378A1283C CRC64;

Query Match 8.9%; Score 165.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 1.9e-07;
Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRLQDNGQCHADAKCVDLHFQDTTVGVFHLRSLP 64
```

```
Db 135 GRYKCE-----VIEGL-----EDDTAVVALDLQGV---VFPYFRL 167
QY 65 GOYKLTEDKAREACANEATMATYNOLSYQXKAKYHLCAGWLETGRVAYPTAFASQNG 124
Db 168 GRYNLNFHEAQACLDQDAVIASFQDLYDAWRGLDWCNAGWLSGDSGVQYPTTKPRECG 227
QY 125 --SGVVGVIVDYGPRPNKSEMMDVFCY 148
Db 228 GQNTVPGVRNYGFWDKSKSYDVFCF 253

RESULT 10
PLK_CHICK
ID PLK_CHICK STANDARD; PRT; 355 AA.
AC P07354;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).
GN CRTLL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYONIC STERNAL CARTILAGE;
RX MEDLINE=8623315; PubMed=3459154;
RA Deak F., Kiss I., Sparks K.J., Argaves W.S., Hampikian G.,
  Goetinck P.F.;
RT "Complete amino acid sequence of chicken cartilage link protein
  deduced from cDNA clones.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3766-3770(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87317659; PubMed=3476955;
RA Kiss I., Deak F., Mestric S., Delius H., Soos J., Dekany K.,
  Argaves W.S., Sparks K.J., Goetinck P.F.;
RT "Structure of the chicken link protein gene: exons correlate with the
  protein domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6399-6403(1987).
CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
  HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35038; AAA48941.1; -
CC EMBL; M35035; AAA48941.1; JOINED.
CC EMBL; M35036; AAA48941.1; JOINED.
CC EMBL; M35037; AAA48941.1; JOINED.
CC EMBL; M13212; AAA48940.1; -
CC PIR; A24881; LKCH.
CC PIR; A28305; A28305.
CC HSP; P98066; ITS.
CC INTERPRO; IPR000538; -
CC INTERPRO; IPR003006; -
CC PFAM; PF00193; Xlink; 2.
CC PFAM; PF00047; ig; 1.
CC PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT PROPEP 1 9
FT CHAIN 10 355 PROTEOGLYCAN LINK PROTEIN.
FT DOMAIN 54 147 IG-LIKE V-TYPE DOMAIN.
```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X17405; CAA35462.1; -;
DR EMBL; U43328; AAA85216.1; -;
DR PIR; S14914; LKHU.
DR PIR; A36308; A36308.
DR HSSP; P98056; 1TSG.
DR MIM; 115435; -;
DR INTERPRO; IPR000538; -;
DR INTERPRO; IPR003006; -;
DR PFAM; PF00193; Xlink; 2.
DR PFAM; PF00047; ig; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
KW PROPEP 1 15
FT CHAIN 16 354
FT DOMAIN 54 146
FT DOMAIN 176 253
FT DOMAIN 274 350
FT DISULFID 61 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 349
FT DISULFID 304 325
FT CARBOHYD 21 21
FT CARBOHYD 56 56
SQ SEQUENCE 354 AA; 40165 MW; 315C96EC3AC2626A CRC64;
PROTEOGLYCAN LINK PROTEIN.
IG-LIKE V-TYPE DOMAIN.
LINK 1.
LINK 2.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
N-LINKED (GLCNAC. -.) (POTENTIAL).
N-LINKED (GLCNAC. -.) (POTENTIAL).
SEQUENCE 354 AA; 40165 MW; 315C96EC3AC2626A CRC64;

Query Match 8.8%; Score 163.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 2.8e-07;
Matches 44; Conservative 21; Mismatches 52; Indels 29; Gaps 4;
4;

QY 5 GKHKCECKSHYVGDGNGCEPEQLPIDRCLQDNGCHADAKVDLHFQDTTVGVFHLRSP 64
Db 135 GRYKCE-----VIEG-----EDTVVVALDIQGV---VFYFPR 167
QY 65 GOYKFTDKAREACANEAAATMATYNQLSYXQAKYHLCASAGWLETRVAYPTAFASQNC 124
Db 168 GRYLNFHEAQACLDQDAVIASFQLYDAWRGGLWCNAGWSDGSVQYPTTKPRECG 227
QY 125 --SGVVGIVDYGPRNKKSEMMDVFCY 148
Db 228 GQNTVPGVGRNYGFWDKRSYDVFCF 253

RESULT 13
PLK_PIG
ID PLK_PIG STANDARD; PRT; 354 AA.
AC P10859;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).
GN CRTLL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=LARYNGEAL CARTILAGE CHONDROCYTES;
RX MEDLINE=89293937; PubMed=2738916;
RT Perkins S.J., Nealis A.S., Dudhja J., Hardingham T.E.;
RA "Immunoglobulin fold and tandem repeat structures in proteoglycan N-
terminal domains and link protein".
RL J. Mol. Biol. 206:737-753(1989).

```
CC CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
CC CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; Y00165; CAA68358.1; -
DR DR PIR; S04243; S04243.
DR DR HSSP; P98066; ITSG.
DR DR INTERPRO; IPR000538; -
DR DR INTERPRO; IPR003006; -
DR DR PFAM; PF00193; Xlink; 2.
DR DR PFAM; PF00047; Ig; 1.
DR DR PROSITE; PS01241; LINK; 2.
KW KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT FT PROPEP 1 9
FT CHAIN 10 354 PROTEOGLYCAN LINK PROTEIN.
FT DOMAIN 54 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 350 LINK 2.
FT DISULFID 61 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 349 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 354 AA; 40260 MW; 68FFB9DE51ABCC1 CRC64;

Query Match 8.7%; Score 161.5; DB 1; Length 354;
Best Local Similarity 29.5%; Pred No. 4.3e-07;
Matches 43; Conservativity 22; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHKCEKSHYVGDGLNCEPEQLPDRCLQDNGQCHADAKCVLDLHFQDTTGVFHLRSLP 64
DQ 135 GRYKCE-----VTEGL-----EDDFAVVALDEGV---VFYFFPRL 167
QY 65 GOYKLTEDKAREACANEATMATYNOLSYQKAKYHLCSAGWLETGRVAYPTAFASQNG 124
DQ 168 GRYNLNFEHQACLDQDAVIAFDQLYDNRGGLDWCNAGWLSQGSVOYPTTKPREPCG 227
QY 125 --SGVGVGDYGPKNKSEMDVFCY 148
DQ 228 GQNTVPGVRNYGEWDXDKSRDYVFCF 253

RESULT 14
PGCA_HUMAN STANDARD; PRT; 2415 AA.
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP) (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 1).
GN AGC1 OR CSPG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE OF 1-2162; 2201-2329 AND 2391-2415 FROM N.A.
RC TISSUE=CHONDROCYTES;
RX MEDLINE=91093289; PubMed=1985970;
RA Doege K.J., Sasaki M., Kimura T., Yamada Y.;
```

```
RT RT *Complete coding sequence and deduced primary structure of the human
RT RT cartilage large aggregating proteoglycan, aggrecan. Human-specific
RT RT repeats, and additional alternatively spliced forms.";
RL J. Biol. Chem. 266:894-902(1991).
RN [2]
RP SEQUENCE OF 1778-2162 AND 2201-2415 FROM N.A.
RC TISSUE=CHONDROCYTES;
RA Baldwin C.T., Reginato A.M., Prockop D.J.;
RA Dudhia J., Hardingham T.E.;
RL Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1936-2415 FROM N.A.
RX MEDLINE=89380154; PubMed=2789216;
RA Balgwin C.T., Reginato A.M., Prockop D.J.;
RT RT *A new epidermal growth factor-like domain in the human core protein
RT RT for the large cartilage-specific proteoglycan. Evidence for
RT RT alternative splicing of the domain.";
RL J. Biol. Chem. 264:15747-15750(1989).
RN [4]
RP SEQUENCE OF 764-864 FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE=95128522; PubMed=7827755;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT RT *Length variation in the keratan sulfate domain of mammalian
RT RT aggrecan.";
RL Matrix Biol. 14:323-328(1994).
CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC CC SIMILARITY).
CC CC -!- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST THREE FORMS OF AGGREGAN
CC CC TRANSCRIPTS GENERATED BY ALTERNATIVE EXON USAGE. THE SHORTEST AND
CC CC ALSO THE MOST PREVALENT FORM, LACKS BOTH THE EGF-LIKE DOMAIN, AND
CC CC SUSHI DOMAIN.
CC CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC CC AND G3.
CC CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; M55172; AAA62824.1; -
DR DR EMBL; J05062; AAA35726.1; -
DR DR EMBL; X17406; CAA35463.1; -
DR DR EMBL; S74659; AAC60643.1; -
DR DR PIR; S08042; S08042.
DR DR HSSP; P98066; ITSG.
DR DR MIM; 155760; -
DR DR INTERPRO; IPR000436; -
DR DR INTERPRO; IPR000495; -
DR DR INTERPRO; IPR000538; -
DR DR INTERPRO; IPR000561; -
DR DR INTERPRO; IPR001304; -
DR DR PFAM; PF00008; EGF; 1.
DR DR PFAM; PF00193; Xlink; 4.
```

DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS01241; LINK; 3.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
FT SIGNAL 1 19
FT CHAIN 20 2415
FT DOMAIN 44 140
FT DOMAIN 170 247
FT DOMAIN 268 349
FT DOMAIN 495 572
FT DOMAIN 593 673
FT DOMAIN 2164 2199
FT DOMAIN 2201 2327
FT REPEAT 2331 2389
FT DOMAIN 48 141
FT DOMAIN 152 247
FT DOMAIN 253 349
FT DOMAIN 477 571
FT DOMAIN 578 672
FT DOMAIN 676 848
FT DOMAIN 772 844
FT DOMAIN 851 1497
FT DOMAIN 941 1497
FT DOMAIN 1498 2162
FT DOMAIN 2163 2415
FT DISULFID 51 133
FT DISULFID 175 246
FT DISULFID 199 220
FT DISULFID 273 348
FT DISULFID 297 318
FT DISULFID 500 571
FT DISULFID 524 545
FT DISULFID 598 672
FT DISULFID 621 642
FT DISULFID 2168 2178
FT DISULFID 2173 2187
FT DISULFID 2189 2198
FT DISULFID 2205 2216
FT DISULFID 2233 2325
FT DISULFID 2301 2317
FT DISULFID 2332 2337
FT DISULFID 2361 2388
FT CARBOHYD 126 126
FT CARBOHYD 239 239
FT CARBOHYD 333 333
FT CARBOHYD 387 387
FT CARBOHYD 434 434
FT CARBOHYD 602 602
FT CARBOHYD 657 657
FT CARBOHYD 737 737
FT CARBOHYD 1898 1898
FT VARSPIC 2163 2200
FT VARSPIC 2330 2390
FT CONFLICT 766 766
FT CONFLICT 847 847
FT CONFLICT 1928 1928
FT CONFLICT 1964 1964
FT CONFLICT 2070 2070
FT CONFLICT 2391 2391
SQ SEQUENCE 2415 AA; 250191 MW; 12889375E1B98C5B6 CRC64;
Query Match 8.7%; Score 161.5; DB 1; Length 2415;
Best Local Similarity 27.9%; Pred. No. 4.6e-06;
Matches 50; Conservative 22; Mismatches 64; Indels 43; Gaps 6;
47 DLHFQDFTVG-----VFHLRSLPGQYKLTDFDKAREACANEATMATYNLSYXQA 97

Db 460 DLVVQVTAVPQGPPLPGVVFHYRPGTPRYSLSLTFEEAQACPGTGAIVASPEQLQAAYEA 519
Qy 98 KYHLCSAGWLETGRVAVPTAFASQNC---GSGVVGIVDYGPRNPKSEMMDVFCY----- 148
Db 520 GYEOCDAGLRDQVRPIVSPRTPCVGDKDSSPGVTVYGRFS-TETVDYVCFVDRLEG 578
Qy 149 -----RWKDV-----NCTXXKVGVDGFS-----YSGNLLQVLSMFFSLT 183
Db 579 EVFFATRLQEQTFQEALEFCFESHNATATTGQLYAANSRGLDKCYAGWLADSLRYPITV 637
RESULT 15
PCGN_MOUSE
ID PCGN_MOUSE STANDARD; PRT; 1268 AA.
AC P55066;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEUROCAN CORE PROTEIN PRECURSOR.
GN CSPG3 OR NCAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=BRAIN;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RA Faessler R.;
RL "Structure and chromosomal localization of the mouse neurocan gene.";
Genomics 28:405-410(1995).
CC -!- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM
AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
ACID.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CUSHI (SCR) REPEAT.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X84727; CAA59216.1; -
DR HSSP; P00740; LIXA.
DR MGD; MGI:104694; CSPG3.
DR INTERPRO; IPR000152; -
DR INTERPRO; IPR000436; -
DR INTERPRO; IPR000538; -
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR001304; -
DR INTERPRO; IPR001881; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00008; EGF; 2.
DR PFAM; PF00193; Xlink; 2.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.

Search completed: April 4, 2001, 13:06:41
Job time: 140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:57 ; Search time 17.49 Seconds

(without alignments)
1370.436 Million cell updates/sec

Title: US-09-466-778-11

Perfect score: 1857
Sequence: 1 MTGPGKHCKESHVGDGL.....ALAAYSYFRINRKTIGFXHF 353

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1468	79.1	1069	T42681	hypothetical prote
2	264	14.2	275	JC6506	tumor necrosis fac
3	264	14.2	276	A47290	TSG-6 homolog P54
4	262	14.1	277	A41735	hyaluronate-bindin
5	175	9.4	2109	I50421	aggrecan precursor
6	171	9.2	2124	A28452	proteoglycan core
7	170	9.2	2327	T42630	aggrecan - bovine
8	169.5	9.1	408	LKRT2	proteoglycan link
9	166.5	9.0	354	S42938	proteoglycan link
10	165.5	8.9	355	LKCH	proteoglycan link
11	165	8.9	2132	A55182	aggrecan precursor
12	163.5	8.8	354	LKHU	proteoglycan link
13	161.5	8.7	354	S04283	proteoglycan link
14	161.5	8.7	2316	A39086	aggrecan precursor
15	160	8.6	370	S29139	aggrecan - pig (fr
16	158.5	8.5	1268	S52781	neurocan - mouse
17	157.5	8.5	371	A53908	brevican precursor
18	157.5	8.5	883	S49126	brevican precursor
19	157.5	8.5	1257	S28764	neurocan precursor
20	155.5	8.4	912	A54423	brevican precursor
21	154.5	8.3	862	S43922	versican - pig-tai
22	152.5	8.2	3562	A47171	chondroitin sulfat
23	151.5	8.2	1643	T14274	versican precursor
24	151.5	8.2	3381	T42389	versican precursor
25	150.5	8.1	883	S57653	brevican precursor
26	150.5	8.1	2409	A60979	versican precursor
27	148	8.0	2397	A55535	versican precursor
28	147	7.9	378	T146268	brevican precursor
29	139.5	7.5	1340	A39808	proteoglycan core

30	122	6.6	113	2	A55885	chondroitin sulfat
31	111.5	6.0	362	2	A35616	T-cell surface gly
32	109.5	5.9	358	2	A34907	plasma membrane gl
33	109.5	5.9	363	2	A37009	CD44 homolog membr
34	108	5.8	395	2	I77371	CD44R5 - human
35	106	5.7	359	2	S24240	lymphocyte surface
36	105	5.7	1584	2	T22674	hypothetical prote
37	104	5.6	3507	2	T34513	hypothetical prote
38	103.5	5.6	365	2	A34424	CD44 membrane gly
39	100	5.4	351	2	S45305	CD44 antigen precu
40	99.5	5.4	728	2	I50719	C-Delta-1 - chicke
41	99.5	5.4	3623	2	T09456	intrinsic factor-B
42	99	5.3	601	2	T22025	hypothetical prote
43	99	5.3	2907	2	A57278	fibrillin-2 precu
44	98	5.3	294	2	A32377	lymphocyte surface
45	98	5.3	361	2	G02251	cell surface glyco

ALIGNMENTS

RESULT 1

T42681
hypothetical protein DKF2p434E0321.1 - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T42681
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A:Reference number: Z22233
A:Accession: T42681
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-870:871-1069 <AAA>
A:Cross-references: EMBL:AL133021
A:Experimental source: adult testis; clone DKF2p434E0321
A:Note: the cDNA sequence contains a +1 frameshift near codon 870
C:Genetics:
A:Note: DKF2p434E0321.1

Query Match	79.1%	Score 1468;	DB 2;	Length 1069;
Best Local Similarity	80.5%;	Pred. No. 1.1e-117;		
Matches 284;	Conservative 4;	Mismatches 29;	Indels 36;	Gaps 2;
QY	1	MTGPGKHCKESHVYGDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTVGVPHL	60	
Db	697	MTGPGKHCKESHVYGDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTVGVPHL	756	
QY	61	RSPGLQYKLTDFDKAREACANEATMATYNQLSYQAKYHLCAGWLETGRVAYPTAFAS	120	
Db	757	RSPGLQYKLTDFDKAREACANEATMATYNQLSYQAKYHLCAGWLETGRVAYPTAFAS	816	
QY	121	QNCSSGVVGIYDYGPRNKSEMDVFCYRKMDVNCYTKVGYVDGFSYSGNLQLVLMSEF	180	
Db	817	QNCSSGVVGIYDYGPRNKSEMDVFCYRKMG-----SAGLFQQLSSRP	860	
QY	181	SLTNFLFEVLAYSNSARGRAFLHLDLSIRGTLFXPQNSGLGENETLSGRDIEHHLAN	240	
Db	861	CIS-----RTPDDLIRGTLFVPQNSGLGENETLSGRDIEHHLAN	900	
QY	241	VSMEFFNDLVNGTTLQTRLGSKLLITDRODPLHPTTRCVGDRDTEWDICASNGITHVI	300	
Db	901	VSMEFFNDLVNGTTLQTRLGSKLLITASQDPLQPTRETFVDGRRAILQWDIFASNGIITHVI	960	
QY	301	SRXLKAPPAPVTLXHTXGLXGIFXXIILVTGAVALAAYSFRINRKTIGFXHF	353	
Db	961	SRPLKAPPAPVTLTHTGLGAGIFAILVLTGAVALAAYSFRINRRTIGFOHF	1013	

RESULT 2

JC6506
tumor necrosis factor stimulated gene-6 protein - mouse

C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: J06506
R:Fueloep, C.; Kamath, R.V.; Li, Y.; Otto, J.M.; Salustri, A.; Olsen, B.R.; Glant, T.T.;
Gene 202, 95-102, 1997
A:Title: Coding sequence, exon-intron structure and chromosomal localization of murine T
A:Reference number: J06506; MUID:98087423
A:Accession: J06506
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-275 <FUE>
A:Cross-references: GB:U83903; NID:q2062474; PID:AA53527.1; PID:q2062475

	Query Match	14.2%	Score 264;	DB 2;	Length 275;	
	Best Local Similarity	43.3%;	Pred. No.	3.le-15;		
	Matches 45; Conservative	16;	Mismatches	43;	Indels	Gaps 0;
Qy	52 DTTVGVFHLRSLPGYKLTFDKAREACANEAAATWATYNOLSYXOKAKYHLCAGWLGTGR	lll : : : :	: : : :	:	: : : :	:
Db	32 EQAAGVYTHREARAGRYKLTAEEAVGCEFEFEGRLUATYKOLEAARKIGFHVCAGGNMAKR	91 : : : :	: : : :	:	:	:
Qy	112 VAYPTAFASQCGSGGVGIVDYGPRPNKSEMDWDVFCYRMKDUNC	155 : : : :	: : : :	:	:	:
Db	92 VGYPVPKGPNCGFKGTIIDYGLRLNRSERWDAYCYNPHAKEC	135 : : : :	: : : :	:	:	:

RESULT 3
A47290
TSG-6 homolog PS4 - rabbit
N:Alternate names: probable cell growth and differentiation protein, 12K; serum-inducible protein 12K; TSG-6
C:Species: *Oryctolagus cuniculus* (domestic rabbit)
C:Date: 03-Mar-1994 #sequence_revision 07-Jul-1995 #text_change 08-Oct-1999
C:Accession: A48055; A47290

J. Biol. Chem. 268, 21453, 1993
A:Reference number: A48055; MUID:94012707
A:Accession: A48055
A:Molecule type: mRNA
A:Residues: 1-276 <FEN>
A:Cross-references: GB:M86381; NID:g387866; PIDN:AAA03342.1; PID:g387867
A:Note: this report replaces the sequence from reference A47290 and shares no similarity
R:Feng, P.; Liau, G.
J. Biol. Chem. 268, 9387-9392, 1993
A:Title: Identification of a novel serum and growth factor-inducible gene in vascular smooth muscle cells
A:Reference number: A47290; MUID:53252803
A:Accession: A47290
A:Molecule type: mRNA
A:Residues: 'MEGNNRKSQLTEDVALDGTGTCRNSFVLTVVRNGLYALVSLGNRRRPQVTVQRTGPNPSSRHSCAREPALQ' (1-100)
A:Note: sequence extracted from vascular smooth muscle cells
A:Note: the sequence reported here was found to be artifactual, a hybrid molecule of two
C:Superfamily: C1r/C1s repeat homology; link protein repeat homology
F:53-128/Domain: link protein repeat homology <LNK>
F:135-244/Domain: C1r/C1s repeat homology <G1R2>

[illegible]

RESULT 4 •
A41735
hyaluronate-binding protein TSG-6 precursor - human

C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C:Accession: A41735; D53642
R:Lee, T.H.; Wisniewski, H.G.; Vilcek, J.
J. Cell Biol. 116, 545-557, 1992
A:Title: A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member
A:Reference number: A41735; MUID:92112993
A:Accession: A41735
A:Molecule type: mRNA
A:Residues: 1-277 <LEE>
A:Cross-references: GB:M31165; NID:g339994; PIDN:AAB00792.1; PID:gl332377
A:Experimental source: FS-4 fibroblast
A:Note: sequence extracted from NCBI backbone (NCBIN:76828, NCBI:P:76833)
R:Wisniewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilcek, J.
Biochemistry 33, 7423-7429, 1994
A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable co
A:Reference number: A53642; MUID:9421799
A:Accession: D53642
A>Status: preliminary
A:Molecule type: protein
A:Residues: 18-27 <WIS>
C:Superfamily: C1r/C1s repeat homology; link protein repeat homology
C:Keywords: glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-277/Product: hyaluronate-binding protein TSG-6 #status predicted <MAT>
F:53-128/Domain: link protein repeat homology <LINK>
F:135-244/Domain: C1r/C1s repeat homology <C1r>
F:118,258/Binding site: carbohydrate (Asn) #status predicted

	Query Match	14.1%	Score 262;	DB 2;	Length 277;
	Best Local Similarity	43.3%;	Pred. No. 4.7e-15;		
	Matches 4;	Conservative	16;	Mismatches 43;	Indels 0; caps 0;
Qy	52 DTTVGVFHLRSLPGQYLTFDKAREACANEAAATWATYNQLSYXOKAKYHLCSGAGLWTGR	111			
	: :				
Dd	32 ERAAGVYTHREARSKYLTYAEAKAVCFEFGHHLYATKQLEAARKIGFHVCAAGMMAKR	91			
	: :				
Qy	112 VAYPTAFASQCNGSGGVGIVDYGPRPNKSENMDWFVCYRMKDVCN	155			
	: :				
Dd	92 VGPIPLVPGNCGFGKTGIIDYGIRLNRSERWDAYCNPNAKEC	135			
	: :				

RESULT 5
I50421
aggrecan precursor - chicken
N:Alternate names: cartilage chondroitin sulfate proteoglycan core protein
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000
C:Accession: I50421; S37356; S273796; A25442; A32002; I50216; A37072; B37072
R:Li, H.; Schwartz, N.B.; Vertel, B.M.
J. Biol. Chem. 268, 23504-23511, 1993
A:Title: cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein
A:Reference number: A48884; MUID:94043149
A:Accession: I50421
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Residues: 1-2109 <LIX>
A:Cross-references: GB:I21913; NID:9416133; PIDN:AAB19128.1; PID:9416134
R:Chandrasekaran, L.; Tanzer, M.L.
Biochem. J. 296, 885-887, 1993
A:Reference number: S39796; MUID:94107258
A:Contents: annotation; erratum
A:Accession: S39796
A:Molecule type: mRNA
A:Residues: 1-361, 'DL', 364-600, 'R', 602-999, 'R', 1001-1028, 'P', 1030-1250, 'D', 1252-1602,
A:Cross-references: GB:M88101
R:Chandrasekaran, L.; Tanzer, M.L.
Biochem. J. 288, 903-910, 1992
A:Title: Molecular cloning of chicken aggregan. Structural analyses.
A:Reference number: S27356; MUID:9311968
A:Accession: S27356
A:Molecule type: mRNA

A:Residues: 1-361,'DL',364-600,'R',602-999,'R',1001-1028,'P',1030-1250,'D',1252-1549,'T'
A:Cross-references: EMBL:M88101
R:Sat, S.; Tanaka, T.; Koshier, R.A.; Tanzer, M.L.
Proc. Natl. Acad. Sci. U.S.A. 83, 5081-5085, 1986
A:Title: Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan
A:Reference number: A25442; MUID:86259736
A:Accession: A25442
A:Molecule type: mRNA
A:Residues: 1693-1795,'G',1797-1855,1894-2109 <SAI>
A:Cross-references: GB:M13993; NID:g211654; PIDN:AAA48720.1; PID:g211655
A:Experimental source: sternal cartilage
R:Tanaka, T.; Har-El, R.; Tanzer, M.L.
J. Biol. Chem. 263, 15831-15835, 1988
A:Title: Partial structure of the gene for chicken cartilage proteoglycan core protein.
A:Reference number: A32002; MUID:89008500
A:Accession: A32002
A:Molecule type: DNA
A:Residues: 1893-1987,'S',1989-2022 <TAN>
A:Note: the authors translated the codon TCC for residue 1787 as Phe
R:Krueger, R.C.
J. Biol. Chem. 265, 12088-12097, 1990
A:Title: Chick cartilage chondroitin sulfate proteoglycan core protein: II. Nucleotide sequence
A:Reference number: I50216; MUID:90307744
A:Accession: I50216
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'PA',1044-1559 <KRU>
A:Cross-references: GB:M38187; NID:g211685; PIDN:AAA48731.1; PID:g555441
R:Krueger Jr., R.C.; Fields, T.A.; Hildreth IV, J.; Schwartz, N.B.
J. Biol. Chem. 265, 12075-12087, 1990
A:Title: Chick cartilage chondroitin sulfate proteoglycan core protein. I. Generation and
A:Reference number: A37072; MUID:90307743
A:Accession: A37072
A:Molecule type: protein
A:Residues: 998-1015,'X',1017-1019,'X',1021-1023 <KR2>
A:Note: amino end of 86k core peptide CS-A
A:Accession: B37072
A:Molecule type: protein
A:Residues: 1247-1250,'D',1252-1272,'X',1274-1275 <KR3>
A:Note: amino end of 75k core peptide CS-B
A:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EC
C:Keywords: alternative splicing
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-2109/Product: aggrecan #status predicted <MAT>
F:44-131/Domain: immunoglobulin homology <IMM>
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-346/Domain: link protein repeat homology <LNK2>
F:537-614/Domain: link protein repeat homology <LNK3>
F:635-716/Domain: link protein repeat homology <LNK4>
F:1859-1890/Domain: BGF homology <EGF>
F:1897-2017/Domain: C-type lectin homology <LCH>
F:2024-2080/Domain: complement factor H repeat homology <FHD>
Query Match 9.4%; Score 175; DB 1; Length 2109;
Best Local Similarity 23.4%; Pred. No. 1.7e-06;
Matches 75; Conservative 31; Mismatches 116; Indels 98; Gaps 12;
QY 10 ECKSHYVDGLNCEPEQLPIDRCLDNGCQCHA-----DAKCVDLHFDQDTVG 56
DB 194 QCDAGLADQVRYPIIHLPRERCYGDKDFPGVRYTGVRETDETYDYCYAEQMQRK--- 250
QY 57 VFHLRSLPGQKLTDFDKAREACANEATMATYNOLSYXOKAKYHLCAGWLETGRVAYPT 116
DB 251 VFYATSP---EKFFQFAFDKCHSLGARLATGELYLAWKDMDCSAGWLADRSVYPI 307
QY 117 AFASQNCQGVGI--VDYGRP-----PNKSEMMDVFCYRMKDYNV-----TXKVG-YV 162
DB 308 SRARNPCGNLGVRYTINLPANQGTGPHPSRYDAICYSGDDPEALVPGLTFDEVGTGL 367
QY 163 GDGFSY-----SCNLLQVLMSPFSLNLF 186
DB 368 GSAFTIQVTVQTEVELPLPRNVTEEARGSATLEPMEITATATELYEAFTVLPDL--FA 425

QY 187 TEVLAYSNSARGRAFLEHLTDL-----SIRGTLFXPQNSGLGE----- 225
DB 426 TSVTVETASPRENTVTREITGWAVEPTVTSVGTAF---TTGMAEVSVEEIAVTA 482
QY 226 ---NETLSGRDIEHHLANVS 242
DB 483 TPGLESAPFTIEDHLVQVT 502
RESULT 6
A28452
proteoglycan core protein precursor, cartilage - rat
N:Alternate names: aggrecan
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 21-Jan-2000
C:Accession: A28452; A28453; A28454; A28455; A28456
R:Doerge, K.; Sasaki, M.; Horiigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 262, 17757-17767, 1987
A:Title: Complete primary structure of the rat cartilage proteoglycan core protein de
A:Reference number: A28453; MUID:88087070
A:Accession: A28453
A:Molecule type: mRNA
A:Residues: 1-2124 <DOE>
R:Doerge, K.; Sasaki, M.; Horiigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 263, 10040a, 1988
A:Reference number: A30069
A:Contents: annotation; revision to residue 698
R:Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 261, 8108-8111, 1986
A:Title: Partial cDNA sequence encoding a globular domain at the C terminus of the ra
A:Reference number: A28453; MUID:86250698
A:Accession: A28453
A:Molecule type: mRNA
A:Residues: 1856-2124 <DO2>
A:Cross-references: GB:M13518; NID:g206104; PIDN:AAA41836.1; PID:g206105
R:Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 262, 17768-17778, 1987
A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-t
A:Reference number: A28453; MUID:88087071
A:Accession: A28453
A:Molecule type: protein
A:Residues: 20-37,'W',39-60,'E',62-64,'X',66-69;70-83;84,89-148,'L',150-238,'S',240,'
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2124/Product: proteoglycan core protein #status predicted <MAT>
F:44-135/Domain: immunoglobulin homology <IMM>
F:170-247/Domain: link protein repeat homology <LNK1>
F:268-349/Domain: link protein repeat homology <LNK2>
F:504-581/Domain: link protein repeat homology <LNK3>
F:602-683/Domain: link protein repeat homology <LNK4>
F:1914-2034/Domain: C-type lectin homology <LCH>
F:2041-2097/Domain: complement factor H repeat homology <FHD>
F:126,239,333,387,611,667,1842/Binding site: carbohydrate (Asn) (covalent) #status pr
Query Match 9.2%; Score 171; DB 2; Length 2124;
Best Local Similarity 28.8%; Pred. No. 3.7e-06;
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;
QY 57 VFHLRSLPGQKLTDFDKAREACANEATMATYNOLSYXOKAKYHLCAGWLETGRVAYPT 116
DB 488 VFHRYGSTRYSITFEAQOACIRTCAAATSPQLOAAVEAGYEQCDAGWLDQTVRYPI 547
QY 117 AFASQNC---GSGVGVIVDYGPRPNKSEMDVFCYRMKDYNVCTXKVGXVGGFSYSGNLL 173
DB 548 VSPRTPCVGDKDSPPGRTYGVRRPS-SETDYVCYVDK-----LEGEVF----- 590
QY 174 QVLMSPFSLTNFLTEVLAYSNSARGRAFLEHLTDLISRTGLTFXPQNSGL 223
DB 591 -----FATQMEQFTFOEAQFCAQNAQT-LASTGQLYAAWSOGL 628

```

RESULT 7
T42630
aggreacan - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Hering, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggreacan: comparative structural analysis
A:Reference number: 222182
A:Accession: T42630
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2327 <HER>
A:Cross-references: EMBL:U76615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
A:Experimental source: articular chondrocytes
C:Superfamily: aggreacan; C-type lectin homology; complement factor H repeat homology; EGF
C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

Query Match 9.2%; Score 170; DB 2; Length 2327;
Best Local Similarity 38.9%; Pred. No. 5e-06;
Matches 37; Conservative 12; Mismatches 42; Indels 4; Gaps 2;

QY 57 VFHLRSLPGQYKLFDRKAREACANEATMATYNOLSYXQKAKYHLCSSAGWLETGRVAYPT 116
Db 488 VFHFRPSSRSLTFEEAKQCLRTGALIASPEQLQAYEAGYEQCDAGWLQDQTVRIPI 547

QY 117 AFASQNC---GSGVGVIVDYGPRPNKSEMDVFCY 148
Db 548 VSPRTPCVGDKSSPGVRTYGVRP-PSETYDVICY 581

RESULT 8
LKRT2
proteoglycan link protein 2 precursor - rat
N:Alternate names: cartilage link protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Aug-1986 #sequence_revision 24-Oct-1997 #text_change 13-Nov-1998
C:Accession: A28654; A24880; A02869
R:Rhodes, C.; Doege, K.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 263, 6063-6067, 1988
A:Title: Alternative splicing generates two different mRNA species for rat link protein.
A:Reference number: A28654; MUID:88198139
A:Accession: A28654
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-408 <RHO>
R:Doege, K.; Hassell, J.R.; Caterson, B.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 83, 3761-3765, 1986
A:Title: Link protein cDNA sequence reveals a tandemly repeated protein structure.
A:Reference number: A24880; MUID:86233314
A:Accession: A24880
A:Molecule type: mRNA
A:Residues: 180-408 <DOE>
R:Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 261, 3519-3535, 1986
A:Title: The primary structure of link protein from rat chondrosarcoma proteoglycan aggregates
A:Reference number: A02869; MUID:86140139
A:Accession: A02869
A:Molecule type: protein
A:Residues: 16-33,'A',89-375,'W',377-408 <NEA>
C:Comment: This protein was extracted from rat chondrosarcoma.
C:Comment: Residues 259-277 and 358-376 (approximately) form epitopes for the species-specific
C:Comment: Link proteins interact with and stabilize aggregates of hyaluronic acid and
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; chondroitin sulfate proteoglycan; duplication; glycoprotein; hyal
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-408/Product: proteoglycan link protein 2 #status experimental <MAT>
F:108-195/Domain: immunoglobulin homology <IMM>
F:230-307/Domain: link protein repeat homology <LNK1>
F:328-404/Domain: link protein repeat homology <LNK2>

```

```

F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:115-193,233-306,259-280,333-403,358-379/Disulfide bonds: #status experimental

Query Match 9.1%; Score 169.5; DB 1; Length 408;
Best Local Similarity 30.8%; Pred. No. 6.4e-07;
Matches 45; Conservative 20; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHLRSP 64
Db 189 GRYKCE-----VIEGL-----EDDTAVVALLQGV---VFYFFPRL 221

QY 65 GOYKLTFRKAREACANEATMATYNOLSYXQKAKYHLCSSAGWLETGRVAYPTAFASQNC 124
Db 222 GRYNLNFHEARQAQLDQDAVIASFDQLYDAWRGGLDWCNAGWLSGDSVQYPIITKPRECG 281

QY 125 --SGVVGVIVDYGPRPNKSEMDVFCY 148
Db 282 GQNTVPGVNRNYGFWDKSKSYDVFCF 307

RESULT 9
S42938
proteoglycan link protein precursor - horse
N:Alternate names: cartilage link protein
C:Species: Equus caballus (domestic horse)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 22-Jun-1999
C:Accession: S42938
R:Dudhia, J.; Platt, D.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42938
A:Accession: S42938
A:Molecule type: mRNA
A:Residues: 1-354 <DUU>
A:Cross-references: EMBL:X78077; NID:g459438; PIDN:CAA54987.1; PID:g459439
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-354/Product: proteoglycan link protein #status predicted <MAT>
F:54-141/Domain: immunoglobulin homology <IMM>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-350/Domain: link protein repeat homology <LNK2>
F:21,56/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:61-139/Disulfide bonds: #status predicted

Query Match 9.0%; Score 166.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 9.7e-07;
Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;

QY 5 GKHKCECKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHLRSP 64
Db 135 GRYKCE-----VIEGL-----EDDTAVVALLQGV---VFYFFPRL 167

QY 65 GOYKLTFRKAREACANEATMATYNOLSYXQKAKYHLCSSAGWLETGRVAYPTAFASQNC 124
Db 168 GRYNLNFHEARQAQLDQDAVIASFDQLYDAWRGGLDWCNAGWLSGDSVQYPIITKPRECG 227

QY 125 --SGVVGVIVDYGPRPNKSEMDVFCY 148
Db 228 GQNTVPGVNRNYGFWDKSKSYDVFCF 253

RESULT 10
LKCH
proteoglycan link protein precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 22-Jun-1999
C:Accession: A28305; A24881; A39097; B39097
R:Kiss, I.; Deak, F.; Mestric, S.; Dellus, H.; Soos, J.; Dekany, K.; Argraves, W.S.;
Proc. Natl. Acad. Sci. U.S.A. 84, 6399-6403, 1987
A:Title: Structure of the chicken link protein gene: exons correlate with the protein
A:Reference number: A28305; MUID:87317659

```

A:Accession: A28305
A:Molecule type: DNA
A:Residues: 1-355 <KIS>
A:Cross-references: GB:M35038; NID:q212264; PIDN:AAA4894.1; PID:g212267
R:Deak, F.; Kiss, I.; Sparks, K.J.; Argraves, W.S.; Hampikian, G.; Goetinck, P.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3766-3770, 1986
A:Title: Complete amino acid sequence of chicken cartilage link protein deduced from cDN
A:Reference number: A24881; MUID:86233315
A:Accession: A24881
A:Molecule type: mRNA
A:Residues: 1-355 <DEA>
A:Cross-references: GB:M31212; NID:q212259; PIDN:AAA4894.1; PID:g212260
A:Experimental source: embryonic sternal cartilage
R:Wu, L.N.Y.; Genge, B.K.; Wuthier, R.E.
J. Biol. Chem. 266, 1187-1194, 1991
A:Title: Association between proteoglycans and matrix vesicles in the extracellular matr
A:Reference number: A39097; MUID:91093230
A:Accession: A39097
A:Molecule type: protein
A:Residues: 40-55, 'X', 57-60, 'X', 62-75, 'X', 77-78, 'X' <WUA>
A:Note: 38k protein, a major component of matrix vesicles; Asn-56 appears to be glycosyl
A:Accession: B39097
A:Molecule type: protein
A:Residues: 40-55, 'X', 57-59 <WU2>
A:Note: 39k protein, a minor component of matrix vesicles
A:Comment: Link proteins stabilize the aggregates of proteoglycan monomers with hyaluron
structures that induce mineralization).
C:Genetics:
A:Introns: 34/1; 159/1; 260/1
A:Note: single copy gene
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-35/Product: link protein #status predicted <MAT>
F:34-142/Domain: immunoglobulin homology <IMM>
F:177-254/Domain: link protein repeat homology <LNK1>
F:275-351/Domain: link protein repeat homology <LNK2>
F:21/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:56/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:61-140,182-253,206-227,280-350,305-326/Disulfide bonds: #status predicted

Query Match 8.9%; Score 165.5; DB 1; Length 355;
Best Local Similarity 29.9%; Pred. No. 1.2e-06;
Matches 44; Conservative 26; Mismatches 46; Indels 31; Gaps 5;
5 GKHKCKSHYVGDGNCPEQLPIDRLQDNGQCHADAKVLDLHFQDTTGVFHLRSP- 63
136 GRKCE-----VIEGL-----EDDTAVVALNLE---GVVFPYSPR 167
64 LGQYKLTDFKAREACANEATMATYNQLSYXQKAKYHLCSSAGWLETGRVAYPTAFASQNC 123
168 LGRYNLFNFEAQACLDQSIILASFDQLYEAWRSGLDWCNAGWLSQVQYPTTKPREPC 227
124 G--SGVVGIVDYGPRNPKSEMDVFCY 148
228 GGRNTVPGVRNRYGFWDKERSYDVFCF 254

RESULT 11
A55182
aggrecan precursor - mouse
C:Species: Mus musculus (house mouse)
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000
C:Accession: A55182; S55329; S50207; S51355; I78532; I58123
R:Walcz, E.; Deak, F.; Erhardt, P.; Coulter, S.N.; Fuelleop, C.; Horvath, P.; Doege, K.J.
Genomics 22, 364-371, 1994
A:Title: Complete coding sequence, deduced primary structure, chromosomal localization,
A:Reference number: A55182; MUID:95104847
A:Accession: A55182
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-2132 <WAL>

A:Cross-references: GB:L07049; NID:g678541; PIDN:AAC37670.1; PID:g191772
R:Watanabe, H.; Gao, L.; Sugiyama, S.; Doege, K.; Kimata, K.; Yamada, Y.
Biochem. J. 308, 433-440, 1995
A:Title: Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene struc
A:Reference number: S55329; MUID:95289972
A:Accession: S55329
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-567, 'G', 569-1728, 'I', 1730-2132 <WAT1>
A:Cross-references: GB:U22901; NID:g886014
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
Biochim. Biophys. Acta 1219, 613-622, 1994
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues
A:Reference number: S50206; MUID:95035091
A:Accession: S50207
A:Molecule type: mRNA
A:Residues: 350-481, 'R', 483-506 <GLU1>
A:Cross-references: EMBL:X80279; NID:g673432
R:Glumoff, V.
submitted to the EMBL Data Library, July 1994
A:Reference number: S51355
A:Accession: S51355
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 350-383, 'CPVMSORERPWAA' <GLU2>
A:Cross-references: EMBL:X80279
R:Watanabe, H.; Kimata, K.; Line, S.; Strong, D.; Gao, L.Y.; Kozak, C.A.; Yamada, Y.
Nature Genet. 7, 154-157, 1994
A:Title: Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the agg
A:Reference number: I58123; MUID:95004579
A:Accession: I78532
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 211-326 <WAT2>
A:Cross-references: GB:S73722; NID:g765215; PIDN:AAB32160.1; PID:g765216
A:Accession: I58123
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 211-240, 'MCTASLRWRVRSFMRHQRNPSRRQPTS', 'AGGWHGAWPPQASSTWPGRAVWTCALAGW'
A:Cross-references: GB:S73720; NID:g765211; PIDN:AAB32159.1; PID:g765212
C:Genetics:
A:Map position: 7
A:Introns: 253/1
A:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: cartilage; extracellular matrix
F:1-19/Domain: signal sequence #status predicted <SIG>
F:44-135/Domain: immunoglobulin homology <IMM>
F:170-247/Domain: link protein repeat homology <LNK1>
F:268-349/Domain: link protein repeat homology <LNK2>
F:504-581/Domain: link protein repeat homology <LNK3>
F:602-683/Domain: link protein repeat homology <LNK4>
F:1922-2042/Domain: C-type lectin homology <LCH>
F:2049-2105/Domain: complement factor H repeat homology <FHD>

Query Match 8.9%; Score 165; DB 1; Length 2132;
Best Local Similarity 28.6%; Pred. No. 1.2e-05;
Matches 49; Conservative 23; Mismatches 66; Indels 32; Gaps 5;
57 VFHLRSPGLQYKLTDFKAREACANEATMATYNQLSYXQKAKYHLCSSAGWLETGRVAYPT 116
488 VFHVRPGSTRYSLTFFEAQACMHTGAIITASPEQLQAAYEAGYEQCDAGWLQDQTVRPI 547
117 AFASQNC---GSGVVGIVDYGPRNPKSEMDVFCFRMKDVNCTXXKVGVDGFSYSGNLL 173
548 VSPRTFCVGDKDSSPGVRYRVRPS-SETDYVCYVDK-----LEGEVF----- 590
174 QVLMSPSLTNFTLEVLAYSNSNARGAFLEHLTDLISIRGTGLFXPQNSGL 223
591 -----FATRLQEQTFQEARAFCAQNT-LASTGQLYAAWSOGL 628

RESULT 12

LKHU
proteoglycan link protein precursor - human
N:Alternate names: cartilage link protein
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 22-Jun-1999
C:Accession: S14914; S04244; S03868; A36308; S14926; S09309; S08041
R:Dudhia, J.; Hardingham, T.E.
Nucleic Acids Res. 18, 1292, 1990
A:Title: The primary structure of human cartilage link protein.
A:Reference number: S14914; MUID:90206798
A:Accession: S14914
A:Molecule type: mRNA
A:Residues: 1-354 <DUD>
A:Cross-references: EMBL:X17405; NID:g463246; PIDN:CAA35462.1; PID:g34378
A:Note: the authors translated the codon GAT for residue 93 as Ala
R:Dudhia, J.; Hardingham, T.E.
J. Mol. Biol. 206, 749-753, 1989
A:Title: Appendix. Isolation and sequence of cDNA clones for pig and human cartilage link protein.
A:Reference number: S04243
A:Accession: S04244
A:Molecule type: mRNA
A:Residues: 223-354 <DUD2>
A:Cross-references: EMBL:Y00166
A:Note: the authors translated the codon CTG for residue 264 as Arg and TTG for residue 264
R:Nuyen, Q.; Murphy, G.; Roughley, P.J.; Mort, J.S.
Biochem. J. 259, 61-67, 1989
A:Title: Degradation of proteoglycan aggregate by a cartilage metalloproteinase. Evidence for a specific cleavage site.
A:Reference number: S03868; MUID:89246328
A:Accession: S03868
A:Molecule type: protein
A:Residues: 16-35 <NGU>
R:Osborne-Lawrence, S.L.; Sinclair, A.K.; Hicks, R.C.; Lacey, S.W.; Eddy Jr., R.L.; Byer
Genomics 8, 562-567, 1990
A:Title: Complete amino acid sequence of human cartilage link protein (CRTLI) deduced from complementary DNA.
A:Reference number: A36308; MUID:91139126
A:Accession: A36308
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-354 <OSB>
R:Dudhia, J.; Hardingham, T.E.
Nucleic Acids Res. 18, 2214, 1990
A:Title: The primary structure of human cartilage link protein.
A:Reference number: S14926; MUID:90245703
A:Accession: S14926
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-92, 'A', 94-354 <NUC>
R:Nuyen, Q.; Mort, J.S.; Roughley, P.J.
Biochem. J. 266, 569-573, 1990
A:Title: Cartilage proteoglycan aggregate is degraded more extensively by cathepsin L than by cathepsin K.
A:Reference number: S09309; MUID:90197639
A:Accession: S09309
A:Molecule type: protein
A:Residues: 16-38, 'X', 40-55, 57-60, 62-65, 'X', 67, 'X', 69-80 <NG2>
C:Genetics:
A:Gene: GDB:CRTL1
A:Cross-references: GDB:125232; OMIM:115435
A:Map position: 5q13-q14
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-354/Product: proteoglycan link protein #status experimental <MAT>
F:54-141/Domain: immunoglobulin homology <IMM>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-350/Domain: link protein repeat homology <LNK2>
F:21, 56/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:61-139/Disulfide bonds: #status predicted

Query Match 8.8%; Score 163.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 1.7e-06;
Matches 44; Conservative 21; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHKCEKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTGVFHLRSPL 64
DB 135 GRYKCE-----VIEGL-----EDTTVVVALDLQGV---VFPFPRLL 167
QY 65 GOYKLTFDKAREACANEATMATYNQLSYXOKAKYHLCISAGWLETGRVAYPTAFASONGC 124
DB 168 GRYNLNFEHAQAACLDQDAVIASFQDLYDAWRGLDWCNAGWLSGVSQVYPTKPREPCG 227
QY 125 --SGVVGIVDYGPRPNKSEMDVFCY 148
DB 228 QQNTVPGVRNYGFWDKDKSRIDVFCF 253
RESULT 13
S04243
proteoglycan link protein precursor - pig
N:Alternate names: cartilage link protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 22-Jun-1999
C:Accession: S04243; I47145
R:Dudhia, J.; Hardingham, T.E.
J. Mol. Biol. 206, 749-753, 1989
A:Title: Appendix. Isolation and sequence of cDNA clones for pig and human cartilage link protein.
A:Reference number: S04243
A:Accession: S04243
A:Molecule type: mRNA
A:Residues: 1-354 <DUD>
A:Cross-references: EMBL:Y00165; NID:g2009; PIDN:CAA68358.1; PID:g2010
R:Perkins, S.J.; Nealis, A.S.; Dudhia, J.; Hardingham, T.E.
J. Mol. Biol. 206, 737-753, 1989
A:Title: Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal
A:Reference number: I47145; MUID:89293837
A:Accession: I47145
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-354 <PER>
A:Cross-references: EMBL:Y00165; NID:g2009; PIDN:CAA68358.1; PID:g2010
C:Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C:Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-354/Product: proteoglycan link protein #status predicted <MAT>
F:54-141/Domain: immunoglobulin homology <IMM>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-350/Domain: link protein repeat homology <LNK2>
F:21, 56/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:61-139/Disulfide bonds: #status predicted
Query Match 8.7%; Score 161.5; DB 1; Length 354;
Best Local Similarity 29.5%; Pred. No. 2.6e-06;
Matches 43; Conservative 22; Mismatches 52; Indels 29; Gaps 4;
QY 5 GKHKCEKSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTGVFHLRSPL 64
DB 135 GRYKCE-----VIEGL-----EDTTVVVALDLQGV---VFPFPRLL 167
QY 65 GOYKLTFDKAREACANEATMATYNQLSYXOKAKYHLCISAGWLETGRVAYPTAFASONGC 124
DB 168 GRYNLNFEHAQAACLDQDAVIASFQDLYDAWRGLDWCNAGWLSGVSQVYPTKPREPCG 227
QY 125 --SGVVGIVDYGPRPNKSEMDVFCY 148
DB 228 QQNTVPGVRNYGFWDKDKSRIDVFCF 253
RESULT 14
A39086
aggrecan precursor - human
N:Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglycan
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jan-2000
C:Accession: A39086; S50206; A43919; S46659; S68689; S68646; B43919; C43919
R:Doerge, K.J.; Sasaki, M.; Kimura, T.; Yamada, Y.

```

Query Match      8.6%; Score 160; DB 2; Length 370;
Best Local Similarity 35.8%; Pred. No. 3.7e-06;
Matches 34; Conservative 16; Mismatches 41; Indels 4; Gaps 2;

Qy   57 VFHLRSPGLGKYLTFDKAREACANEATMATYNOLSYXQAKYHLCISAGWLETGRVAYPT 116
      ||| : : | ||| : || : ||| : : : |||| : ||
Db   138 VFHYRAISXYRTLDFDRAQRACLQNLSAIIAPTEQLQAAYEDGFHQCDAGWLADQTVPYI 197

Qy   117 AFASQC---GSGVVGVIDYGPRPNKSEMDVFICY 148
      : | : ||| : : | : ||| :
Db   198 HTPREGCYGDKEFFGVITYGIR-DTNETYDVYCF 231

```

Search completed: April 4, 2001, 13:04:41
Job time: 44 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:56 ; Search time 13.89 Seconds
(without alignments)
456.360 Million cell updates/sec

Title: US-09-466-778-11
Perfect score: 1857
Sequence: 1 MTGPGKHCKECSHYVGDGL.....ALAAYSYFINKTIGTFXHF 353

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6.COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/PCUS.COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262	14.1	277	1	US-08-024-868-2
2	262	14.1	277	2	US-08-242-097-2
3	262	14.1	277	4	PCT-US96-11995-1
4	167	9.0	326	1	US-08-225-477B-6
5	167	9.0	326	2	PCT-US95-04353-6
6	160	8.6	98	2	US-08-242-097-4
7	157.5	8.5	371	1	US-08-225-477B-8
8	157.5	8.5	371	4	PCT-US95-04353-8
9	157.5	8.5	1257	1	US-08-340-428B-49
10	157	8.5	329	1	US-08-225-477B-3
11	157	8.5	329	4	PCT-US95-04353-3
12	156	8.4	333	1	US-08-225-477B-4
13	156	8.4	333	4	PCT-US95-04353-4
14	155.5	8.4	912	4	PCT-US95-03747-2
15	155.5	8.3	97	2	US-08-242-097-5
16	150.5	8.1	328	1	US-08-225-477B-5
17	150.5	8.1	328	4	PCT-US95-04353-5
18	150.5	8.1	2409	5	5180808-2
19	148	8.0	908	4	PCT-US95-03747-3
20	147	7.9	378	1	US-08-225-477B-9
21	147	7.9	378	4	PCT-US95-04353-9
22	107	5.8	322	2	US-08-892-880-2
23	105	5.7	810	2	US-08-820-170A-34
24	105	5.7	810	3	US-09-055-699-34
25	105	5.7	810	3	US-09-273-565-34
26	103.5	5.6	717	3	US-08-872-855-9
27	99	5.3	735	3	US-09-191-647-9
28	98.5	5.3	363	1	US-07-946-497-7

29	98.5	5.3	363	1	US-08-483-322-7
30	98.5	5.3	363	2	US-08-478-882-7
31	97.5	5.3	339	2	US-08-892-880-3
32	97.5	5.3	503	1	US-07-946-497-2
33	97.5	5.3	503	2	US-08-483-322-2
34	97.5	5.3	503	2	US-08-478-882-2
35	97.5	5.3	1480	3	US-09-191-647-7
36	97.5	5.3	1480	4	PCT-US91-09055-2
37	97	5.2	729	3	US-08-872-855-8
38	96.5	5.2	721	3	US-08-872-855-7
39	96	5.2	361	1	US-07-946-497-6
40	96	5.2	361	2	US-08-483-322-6
41	96	5.2	361	2	US-08-478-882-6
42	94.5	5.1	720	3	US-08-872-855-4
43	92.5	5.0	362	5	5504194-2
44	91.5	4.9	1404	2	US-08-400-159-2
45	91.5	4.9	1404	3	US-08-611-729A-2

ALIGNMENTS

RESULT 1
US-08-024-868-2
; Sequence 2, Application US/08024868
; Patent No. 5386013
; GENERAL INFORMATION:
; APPLICANT: Lee, Tae Ho
; APPLICANT: Wisniewski, Hans Georg
; APPLICANT: Vilcek, Jan
; TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA
; CODING THEREFOR AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/024,868
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/642,312
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Livnat, Shmuel
REGISTRATION NUMBER: 33,949
REFERENCE/DOCKET NUMBER: VILCEK=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-628-5197
TELEFAX: 212-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-024-868-2

Query Match 14.1%; Score 262; DB 1; Length 277;
Best Local Similarity 43.3%; Pred No. 8e-21;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

Db 32 ERAAGVYHREARSGKYKLTAEAKAVCEFGGHLATYKQLEARKIGFHVCAAGWMAKGR 91
QY 112 VAYPTAFASQNCGGVGVVDYGRPNKSEMMDVFCYRMKDVC 155
Db 92 VGYPIVKGPNCGFGTKGIIDYGIRLNRSEWDAYCYNPHAKEC 135

RESULT 2

US-08-242-097-2
; Sequence 2, Application US/08242097
; Patent No. 5846763
; GENERAL INFORMATION:
; APPLICANT: Lee, Tae Ho
; APPLICANT: Wisniewski, Hans Georg
; APPLICANT: Vilcek, Jan
; TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
; TITLE OF INVENTION: Thereof and Uses Thereof
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,097
; FILING DATE: 13-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-628-5197
; TELEFAX: 212-737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-242-097-2

Query Match 14.1%; Score 262; DB 2; Length 277;
Best Local Similarity 43.3%; Pred. No. 8e-21;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;
QY 52 DTTVGVFHLRSLPGQYKLTFDKAREACANEATMATYNQLSYXQAKYHLCISAGWLETGR 111
Db 32 ERAAGVYHREARSGKYKLTAEAKAVCEFGGHLATYKQLEARKIGFHVCAAGWMAKGR 91
QY 112 VAYPTAFASQNCGGVGVVDYGRPNKSEMMDVFCYRMKDVC 155
Db 92 VGYPIVKGPNCGFGTKGIIDYGIRLNRSEWDAYCYNPHAKEC 135

RESULT 3

PCT-US96-11995-1
; Sequence 1, Application PC/TUS9611995
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A Pharmaceutical Composition Containing TSG-6
; TITLE OF INVENTION: Protein for Treating Inflammatory Diseases and Cancer-Related
; TITLE OF INVENTION: Methods of Using same
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11995
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,311
; FILING DATE: 20-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: WISNIEWSKI-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-628-5197
; TELEFAX: 212-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-11995-1

Query Match 14.1%; Score 262; DB 4; Length 277;
Best Local Similarity 43.3%; Pred. No. 8e-21;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;
QY 52 DTTVGVFHLRSLPGQYKLTFDKAREACANEATMATYNQLSYXQAKYHLCISAGWLETGR 111
Db 32 ERAAGVYHREARSGKYKLTAEAKAVCEFGGHLATYKQLEARKIGFHVCAAGWMAKGR 91
QY 112 VAYPTAFASQNCGGVGVVDYGRPNKSEMMDVFCYRMKDVC 155
Db 92 VGYPIVKGPNCGFGTKGIIDYGIRLNRSEWDAYCYNPHAKEC 135

RESULT 4

US-08-225-477B-6
; Sequence 6, Application US/08225477B
; Patent No. 5635370
; GENERAL INFORMATION:
; APPLICANT: Susan Hockfield
; APPLICANT: Diane M. Jaworski
; TITLE OF INVENTION: BEHAB, A Brain Hya-
; TITLE OF INVENTION: Iuronan-Binding Protein
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: St. Onge Steward Johnston & Reens
; STREET: 986 Bedford Street
; CITY: Stamford
; STATE: CT
; COUNTRY: United States
; ZIP: 06905
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Word Processor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,477B
; FILING DATE: April 8, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky

REGISTRATION NUMBER: 32423
REFERENCE/DOCKET NUMBER: 1751-P0004
TELEPHONE: 203-324-6155
TELEFAX: 203-327-1096
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 326 residues
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
DESCRIPTION: functional domains
FRAGMENT TYPE: functional domains
FEATURE:
NAME/KEY: rat link protein
PUBLICATION INFORMATION:
AUTHORS: Doege, K., Hassell, J.R., Ca-
AUTHORS: terson, B., and Yamada, Y.
TITLE: Link protein cDNA sequence reveals a
TITLE: tandemly repeated protein sequence.
JOURNAL: Proc. Natl. Acad. Sci. USA
VOLUME: 83
PAGES: 3761-3765
DATE: 1986
US-08-225-477B-6

Query Match 9.0%; Score 167; DB 1; Length 326;
Best Local Similarity 31.5%; Pred. No. 3.2e-10;
Matches 46; Conservative 21; Mismatches 49; Indels 30; Gaps 5;

QY 5 GKHKCEKSHYVGDLNCEPEQLPIDRCLQDNCQCHADAKCVDLHFQDTTGVGFHLRSL 64
Db 109 GRYKCE-----VIEGL-----EDDTAVVALELQGV---VFFYFPRL 141
QY 65 GOYKLTDFKAREACANEAATMATYNOLSYXOKAKYHLCGAGWLETGRVAYPTAFASQNCG 124
Db 142 GRYNLNPFHARQACLDQDAVIAFDQLYDAWRGGLDWCNAGWLSGDSGVQYPTTKPREPG 201
QY 125 --SGVGVGDYGRPNKSEMDVFCY 148
Db 202 GQNTVPGVRNYG-FWDKDSRYDVFCF 226

RESULT 5
PCT-US95-04353-6
US-08-242-097-4
Sequence 6, Application PC/TUS9504353
GENERAL INFORMATION:
APPLICANT: Susan Hockfield
APPLICANT: Diane M. Jaworski
TITLE OF INVENTION: BEHAB, A Brain Hya-
TITLE OF INVENTION: luronan-Binding Protein
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: St. Onge Steward Johnston & Reens
STREET: 986 Bedford Street
CITY: Stamford
STATE: CT
COUNTRY: United States
ZIP: 06905
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" 1.44 Mb diskette
COMPUTER: IBM PC
OPERATING SYSTEM: MS DOS
SOFTWARE: Word Processor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04353
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,477
FILING DATE: April 8, 1994
ATTORNEY/AGENT INFORMATION:

NAME: Mary M. Krinsky
REGISTRATION NUMBER: 32423
REFERENCE/DOCKET NUMBER: 1751-P0004
TELEPHONE: 203-324-6155
TELEFAX: 203-327-1096
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 326 residues
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
DESCRIPTION: functional domains
FRAGMENT TYPE: functional domains
FEATURE:
NAME/KEY: rat link protein
PUBLICATION INFORMATION:
AUTHORS: Doege, K., Hassell, J.R., Ca-
AUTHORS: terson, B., and Yamada, Y.
TITLE: Link protein cDNA sequence reveals a
TITLE: tandemly repeated protein sequence.
JOURNAL: Proc. Natl. Acad. Sci. USA
VOLUME: 83
PAGES: 3761-3765
DATE: 1986
PCT-US95-04353-6

Query Match 9.0%; Score 167; DB 4; Length 326;
Best Local Similarity 31.5%; Pred. No. 3.2e-10;
Matches 46; Conservative 21; Mismatches 49; Indels 30; Gaps 5;

QY 5 GKHKCEKSHYVGDLNCEPEQLPIDRCLQDNCQCHADAKCVDLHFQDTTGVGFHLRSL 64
Db 109 GRYKCE-----VIEGL-----EDDTAVVALELQGV---VFFYFPRL 141
QY 65 GOYKLTDFKAREACANEAATMATYNOLSYXOKAKYHLCGAGWLETGRVAYPTAFASQNCG 124
Db 142 GRYNLNPFHARQACLDQDAVIAFDQLYDAWRGGLDWCNAGWLSGDSGVQYPTTKPREPG 201
QY 125 --SGVGVGDYGRPNKSEMDVFCY 148
Db 202 GQNTVPGVRNYG-FWDKDSRYDVFCF 226

RESULT 6
US-08-242-097-4
Sequence 4, Application US/08242097
Patent No. 5846763
GENERAL INFORMATION:
APPLICANT: Lee, Tae Ho
APPLICANT: Wisniewski, Hans Georg
APPLICANT: Vlicek, Jan
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
TITLE OF INVENTION: Therefor and Uses Thereof
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,097
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LER26/VILCEK-1B
TELEPHONE: 212-628-5197
TELEFAX: 212-737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 98 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-242-097-4

Query Match 8.6%; Score 160; DB 2; Length 98;

Best Local Similarity 36.2%; Pred. No. 3e-10;
Matches 34; Conservative 17; Mismatches 41; Indels 2; Gaps 1;

QY 57 VFHLRSLGQYKLTDFDKAREACANEATMATYNQLSYXQKAKYHLCAGWLETRVAYPT 116
DB 3 VFYFPLRLGRYLNLFHFAEQACLDQDAVIAFDQLYDAWRGLDWCNAGWLSGDSVQYPI 62

QY 117 AFASQCG--SGVVGIVDYGPRNKSEMDVFCY 148
DB 63 TKPRECGGQTVPGVRNYGFWDKDSRYDVFCF 96

RESULT 7

US-08-225-477B-8
Sequence 8, Application US/08225477B

Patent No. 5633370

GENERAL INFORMATION:

APPLICANT: Susan Hockfield
APPLICANT: Diane M. Jaworski
TITLE OF INVENTION: BEHAB, A Brain Hya-
TITLE OF INVENTION: luronan-Binding Protein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: St. Onge Steward Johnston & Reens
STREET: 986 Bedford Street
CITY: Stamford
STATE: CT
COUNTRY: United States
ZIP: 06905

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" 1.44 Mb diskette
COMPUTER: IBM PC
OPERATING SYSTEM: MS DOS
SOFTWARE: Word Processor
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/225,477B
FILING DATE: April 8, 1994
ATTORNEY/AGENT INFORMATION:

NAME: Mary M. Krinsky
REGISTRATION NUMBER: 32423
REFERENCE/DOCKET NUMBER: 1751-P0004
TELEPHONE: 203-324-6155
TELEFAX: 203-327-1096

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: polypeptide
FRAGMENT TYPE: entire sequence
IMMEDIATE SOURCE: rat brain
FEATURE:

NAME/KEY: rat BEHAB

OTHER INFORMATION: polypeptide encoded by (and set out under) SEQ ID NO 1
OTHER INFORMATION: ed by (and set out under) SEQ ID NO 1
US-08-225-477B-8

Query Match 8.5%; Score 157.5; DB 1; Length 371;

Best Local Similarity 27.9%; Pred. No. 4.3e-09;
Matches 46; Conservative 22; Mismatches 60; Indels 37; Gaps 5;

QY 10 ECKSHVVDGLNCEPEQLPIDRCLQD-----NGQCHADAKCVDLHF 50
DB 201 QCDAGWLSQIVRYPIQNPREACYGDMGYPGVRNYGVVGGDDLDYVCYAE-----DLN- 255

QY 51 QDITVGVFHLRSLGQYKLTDFDKAREACANEATMATYNQLSYXQKAKYHLCAGWLETG 110
DB 256 -----GELFLGAPPG--KLTWEARDYCLERGAQIASTGQLYAAWNGGLDRCSFGWLADG 308

QY 111 RVAYPTAFASQCGSGVVGIVDYGPRN-----KSEMDVFCYR 149

DB 309 SVRYPIITPSQRCGGGLPGVKTFLFPNQTGTFPSKQNRFNVCYCFR 353

RESULT 8

PCT-US95-04353-8

Sequence 8, Application PC/TUS9504353

GENERAL INFORMATION:

APPLICANT: Susan Hockfield
APPLICANT: Diane M. Jaworski
TITLE OF INVENTION: BEHAB, A Brain Hya-
TITLE OF INVENTION: luronan-Binding Protein
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: St. Onge Steward Johnston & Reens
STREET: 986 Bedford Street
CITY: Stamford
STATE: CT
COUNTRY: United States
ZIP: 06905

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" 1.44 Mb diskette
COMPUTER: IBM PC
OPERATING SYSTEM: MS DOS
SOFTWARE: Word Processor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04353

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/225,477

FILING DATE: April 8, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Mary M. Krinsky

REGISTRATION NUMBER: 32423

REFERENCE/DOCKET NUMBER: 1751-P0004

TELEPHONE: 203-324-6155

TELEFAX: 203-327-1096

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 371 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: polypeptide

FRAGMENT TYPE: entire sequence

IMMEDIATE SOURCE: rat brain

FEATURE:

NAME/KEY: rat BEHAB

OTHER INFORMATION: polypeptide encoded by (and set out under) SEQ ID NO 1

OTHER INFORMATION: ed by (and set out under) SEQ ID NO 1

PCT-US95-04353-8

Query Match 8.5%; Score 157.5; DB 4; Length 371;
Best Local Similarity 27.9%; Pred. No. 4.3e-09;
Matches 46; Conservative 22; Mismatches 60; Indels 37; Gaps 5;

QY 10 ECKSHVVDGLNCEPEQLPIDRCLQD-----NGQCHADAKCVDLHF 50
DB 201 QCDAGWLSDDTVRYPIQNPRACYGDMGYPGVNRNYGVVGPDDLVDYCYAE-----DLN- 255

QY 51 QDVTGVFHLRSPGLQYKLTEDKAREACANEATMATYNQLSYKOKAKYHLCAGWLETG 110
DB 256 -----GELFLGAPPG--KLTWEARDYCLERGAQTASTGQLYAAWNGGLDRCSFGWLADG 308

QY 111 RVAYPTAFASQNGSGVGIYDYGPRPN-----KSEMDVFCYR 149
DB 309 SVRYPIIPTSQRCCGGLPGVTLFLFPNQTFPPSKONRFNYICFR 353

RESULT 9
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,428B
; FILING DATE: 14 No. 5648465ember 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-340-428B-49

Query Match 8.5%; Score 157.5; DB 1; Length 1257;
Best Local Similarity 34.2%; Pred. No. 2.8e-08;
Matches 41; Conservative 19; Mismatches 55; Indels 7; Gaps 3;

QY 47 DLHFQDVTGVFHLRSPGLQYKLTEDKAREACANEATMATYNQLSYKOKAKYHLCAGW 106
DB 150 DLVTELVTVGFHYRAARDRYALTFAEQEACHLSSATIAAPRHQAQAFEDGFDCNDAGW 209

QY 107 LETGRVAYPTAFASQNC---GSGVGIYDYGPRPNKSEMDVFCYRBMKDYNCTKVGCVG 163
DB 210 LSDRTVRYPIITQSRPCGYGDRSLPGVRSYG--RRDPQELIDYVCFARE---LGGEVFTVG 265

RESULT 10
US-08-225-477B-3
; Sequence 3, Application US/08225477B
; Patent No. 5635370
; GENERAL INFORMATION:
; APPLICANT: Susan Hockfield
; APPLICANT: Diane M. Jaworski
; TITLE OF INVENTION: BEHAV, A Brain Hya-
; TITLE OF INVENTION: luronan-Binding Protein
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: St. Onge Steward Johnston & Reens
; STREET: 986 Bedford Street
; CITY: Stamford
; STATE: CT
; COUNTRY: United States
; ZIP: 06905

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Word Processor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225.477B
; FILING DATE: April 8, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky
; REGISTRATION NUMBER: 32423
; REFERENCE/DOCKET NUMBER: 1751-P0004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-324-6155
; TELEFAX: 203-327-1096
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 residues
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: polypeptide
; FRAGMENT TYPE: functional domains
; FEATURE:
; NAME/KEY: rat aggregan
; PUBLICATION INFORMATION:
; AUTHORS: Doege, K., Sasaki, M., Hori-
; TITLE: Complete primary structure of the rat
; TITLE: cartilage proteoglycan core protein deduced from
; TITLE: cDNA clones.
; JOURNAL: J. Biol. Chem.
; VOLUME: 262
; PAGES: 17757-17767
; DATE: 1987
US-08-225-477B-3

Query Match 8.5%; Score 157; DB 1; Length 329;
Best Local Similarity 26.8%; Pred. No. 4.1e-09;
Matches 44; Conservative 23; Mismatches 67; Indels 30; Gaps 6;

QY 5 GKHKCEKSHYVGDGLNCEPEQLPIDRC-----LDNGQCHADAKCVDLHF 50
DB 175 GFHQCD--AGWLADQVRYPIHTPREGCYGDKDFPGVTVGYRTNETY-DVYC----F 227

QY 51 QDVTGVFHLRSPGLQYKLTEDKAREACANEATMATYNQLSYKOKAKYHLCAGWLETG 110
DB 228 AEEMEGEFVATSP---EKFTFOEAAANEKRTVGARLATTGQLYLAWQGGMDMCSAGWLADR 284

Search completed: April 4, 2001, 13:04:18
Job time: 22 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 Seconds
(without alignments)
1505.071 Million cell updates/sec

Title: US-09-466-778-11
Perfect score: 1857
Sequence: 1 MTGPKHKCKSHYVDGL.....ALAAVSFRINRKTIGFXHF 353

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB-seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.15.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1757	94.6	897	Q9NRY3	Q9NRY3 homo sapien
2	1468	79.1	1069	Q30F98	Q9uf98 homo sapien
3	662	35.6	2212	Q3072	Q93072 homo sapien
4	645.5	34.8	2570	Q9N15	Q9N15 homo sapien
5	264	14.2	275	11 Q08859	Q08859 mus musculus
6	177	9.5	2109	13 P79787	P79787 gallus gall
7	169.5	9.1	355	11 Q921X7	Q921X7 mus musculus
8	162	8.7	1321	4 Q14594	Q14594 homo sapien
9	161.5	8.7	655	11 Q088564	Q088564 rattus norv
10	161.5	8.7	1290	13 Q9W6E1	Q9W6E1 gallus gall
11	160.5	8.6	2394	6 Q77610	Q77610 bos taurus
12	153	8.2	192	6 Q02817	Q02817 oryctolagus
13	153	8.2	656	6 Q77612	Q77612 bos taurus
14	151.5	8.2	1643	6 Q77611	Q77611 bos taurus
15	151.5	8.2	3381	6 Q77609	Q77609 bos taurus
16	126.5	6.8	103	6 Q9TTB3	Q9TTB3 sus scrofa
17	120.5	6.5	103	6 Q46380	Q46380 oryctolagus
18	110	5.9	396	13 Q9W6S4	Q9W6S4 gallus gall
19	109	5.9	302	4 Q9UNF4	Q9UNF4 homo sapien

ALIGNMENTS

RESULT	1
Q9NRY3	PRELIMINARY; PRT; 897 AA.
AC Q9NRY3	01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000	(TREMREL. 15, Last sequence update)
DT 01-OCT-2000	(TREMREL. 15, Last annotation update)
DE	CD44-LIKE PRECURSOR FELL.
OS	Homo sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Tao Q., Zhang W., Cao X.;
RT	"Molecular cloning and characterization of human FELL sharing homology with CD44."
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF160476; AAF82398.1;
SQ	SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match	94.6%;	Score 1757;	DB 4;	Length 897;
Best Local Similarity	93.8%;	Pred. NO. 8.6e-162;		
Matches 331;	Conservative	3;	Mismatches 19;	Indels 0; Gaps 0;
QY	1	MTGPKHKCKSHYVDGLNCEPEQLPTDRCLODNGOCHADAKCVDLHFQDFTVGVFHL	60	
DB	489	MTGPKHKCKSHYVDGLNCEPEQLPTDRCLODNGOCHADAKCVDLHFQDFTVGVFHL	548	
QY	61	RSLPGYKLTDFKAREACANEATMATYNQLSYXQAKYHLCSAGWLTGRTGVAYPTAFAS	120	
DB	549	RSLPGYKLTDFKAREACANEATMATYNQLSYXQAKYHLCSAGWLTGRTGVAYPTAFAS	608	
QY	121	QNGCGSVGIVDYGPRPNKSEMDVFCYRMKDVCNTXXVGVVGDGFCFSGNLLQVLSFPP	180	
DB	609	QNGCGSVGIVDYGPRPNKSEMDVFCYRMKDVCNTXXVGVVGDGFCFSGNLLQVLSFPP	668	
QY	181	SLTNFTLTVLAYSNSSARGAFLEHLTDLSTGRGLTXPONSGLGENETLSGRDIEHLAN	240	
DB	669	SLTNFTLTVLAYSNSSARGAFLEHLTDLSTGRGLTXPONSGLGENETLSGRDIEHLAN	728	


```
QY 241 VSMFFNDLVNGTTTQTLRLGSKLLITDRQDPLHPTETRCVGDGRDTEWIDICASNGITHVI 300
DB 729 VSMFFNDLVNGTTTQTLRVGSKLLITASQDPLQPTETRFVDGRAILQWIDFASNGIHHVI 788
QY 301 SRXLKAPAPVTLHTXILGXGIFXXIIILVTGAVALAAYSFRINRRTIGFQXHF 353
DB 789 SRPLKAPAPVTLHTGLGAGIFFAILVTGAVALAAYSFRINRRTIGFQHF 841

RESULT 2
Q9UF98 PRELIMINARY; PRT; 1069 AA.
AC Q9UF98;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).
GN DKF2P434E0321.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133021; CAB61358.1;
DR HSP: P98066; ITSG.
DR INTERPRO: IPR000538;
DR INTERPRO: IPR000561;
DR INTERPRO: IPR000782;
DR PFAM: PF00008; EGF; 8.
DR PFAM: PF00193; Xlink; 1.
DR PRINTS: PR00193; Xlink; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 5.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1069 AA; 115666 MW; 04B0950932164D63 CRC64;

Query Match 79.18; Score 1468; DB 4; Length 1069;
Best Local Similarity 80.58; Pred. No. 1.2e-133;
Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 60
DB 697 MTGPGKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 756
QY 61 RSPGQYKLTDFKAREACANEATMAYNQLSYQAKYHLCAGWLETGRVAYPTAFAS 120
DB 757 RSPGQYKLTDFKAREACANEATMAYNQLSYQAKYHLCAGWLETGRVAYPTAFAS 816
QY 121 QNCGSGVGVIVDYGPRPNKSEMDVFCYRMKDVCNCTKKVGVGDGFSYSGNLLQVLMSEFP 180
DB 817 QNCGSGVGVIVDYGPRPNKSEMDVFCYRMKG -----SAGLFQQLSSRP 860
QY 181 SUTNLTFLVAYSNSARGRAFLHLEHTDLSIRGLTFXPQNSGLGENETLSGRDIEHHLA 240
DB 861 CIS-----RTPDDLIRGLTFXPQNSGLGENETLSGRDIEHHLA 900
QY 241 VSMFFNDLVNGTTTQTLRLGSKLLITDRQDPLHPTETRCVGDGRDTEWIDICASNGITHVI 300
DB 901 VSMFFNDLVNGTTTQTLRLGSKLLITASQDPLQPTETRFVDGRAILQWIDFASNGIHHVI 960
QY 301 SRXLKAPAPVTLHTXILGXGIFXXIIILVTGAVALAAYSFRINRRTIGFQXHF 353
DB 961 SRPLKAPAPVTLHTGLGAGIFFAILVTGAVALAAYSFRINRRTIGFQHF 1013
```

```
RESULT 3
Q93072 PRELIMINARY; PRT; 2212 AA.
AC Q93072;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MYELOBLAST KIAA0246 PROTEIN (FRAGMENT).
GN KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.
RL DNA Res. 3:321-329(1996).
DR EMBL: D87433; BAA1337.1;
DR HSP: P98066; ITSG.
DR INTERPRO: IPR000538;
DR INTERPRO: IPR000561;
DR INTERPRO: IPR000782;
DR INTERPRO: IPR001128;
DR PFAM: PF00008; EGF; 13.
DR PFAM: PF00193; Xlink; 1.
DR PRINTS: PR01265; LINKMODULE.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 13.
DR PROSITE: PS01241; LINK; UNKNOWN_1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
KW Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 2212 AA; 237451 MW; 4A95A60504129134 CRC64;
```

```
Query Match 35.68; Score 662; DB 4; Length 2212;
Best Local Similarity 40.28; Pred. No. 4.9e-55;
Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGPGRKHCKECSHYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCVDLHFQDTTVGVFHL 60
DB 1796 TGLNTRRCECHAGYVGDGLQCLESEPPVDRCLGQPPPCHEADACTDLHFQEKRAAGVFL 1855
QY 61 RSPGQYKLTDFKAREACANEATMAYNQLSYQAKYHLCAGWLETGRVAYPTAFAS 120
DB 1856 QATSGPYGLNFSAEARACACQAGVLAAPPQSLAAQQLGFLHLCMLWLANGSTAHPPVFPV 1915
QY 121 QNCGSGVGVIVDYGPRPNKSEMDVFCYRMKDVCNCTKKVGVGDGFSYSGNLLQVLMSEFP 179
DB 1916 ADCGNGRGGVIVDYGPRPNKSEMDVFCYRMKDVCNCTKKVGVGDGFSYSGNLLQVLAAT 1975
QY 180 PSUTNLTFLVAYSNSARGRAFLHLEHTDLSIRGLTFXPQNSGLGENETLSGRDIEHHLA 239
DB 1976 ANFTSYGMLLYANATQGLDFLDDELTYKTLFVPVNEGVDNMTLSGPDLELHAS 2035
QY 240 NVSMFFNDLVNGTTTQTLRLGSKLLITDR---QDPLHPTETRCVGDGRDTEWIDICASNGI 296
DB 2036 NATLLSAN-ASQGLLPAHSGLSLIISDAGPDNNSWAPAGTVVVSRIIVWDIAFNFI 2094
QY 297 THVISRLKAPAPVTLHTXILGXGIFXXIIILVTGAV--ALAAYSFRINRRTIGF 350
DB 2095 IHALASPLAPPQPAVLAPAPPAAGVAVLAAGALLGLVAGALYLARGKPMGF 2150
```


Wed Apr 4 13:03:50 2001

us-09-466-778-11.rag

Seq ID 11, Database: A-Geneseg-36
Page 2

Page 2

[illegible]

New TSG-6 compositions to treat inflammatory diseases and cancer - can be complexed with inter-alpha-inhibitor, to treat auto-immune disease, cancer-related pathology etc. can also be administered as TSG-6 DNA

Claim 2; Page 58-59; 76pp; English.

This protein is the human tumour necrosis factor-stimulated gene-6 protein (TSG-6). It can be used in a claimed pharmaceutical composition for treating inflammatory diseases and disorders, or cancer-related pathologies. The composition or DNA encoding TSG-6 may be used to inhibit tumour metastases, invasive tumour growth or cancer-related pathologies. They may also be used in the treatment of inflammatory conditions including rheumatoid arthritis, systemic lupus erythematosus, and infections such as cachexia, atherosclerosis etc. TSG-6 can also be used to design drugs to treat these disorders.

Sequence 277 AA.

Query Match 14.1%; Score 262; DB 18; Length 277;
Best Local Similarity 43.3%; Pred. No. 1.6e-19;
Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps.

QY 52 DTTGVGFHLRSLPGQYKLFNDAAREACANEATMATYNSLYXQKAKYHLCISAGMLETGR 111
DB 32 eraagvyhrearsgkyllyaeakayceffegghlatykleaarkigfivcaagwmakgr 91

QY 112 VAYTAFASQCSGGVGVVDVGPENKSEMMDVFCYRKMKNVNC 155
DB 92 vgvplvkgpgncgfgktgldygrlnrsewdacycnpshakec 135

RESULT 3

W84087
ID W84087 standard; Protein; 277 AA.
AC
XX W84087;
XX
XX
XX 10-FEB-1999 (first entry)
XX
XX Tumour necrosis factor stimulated gene 6 (TSG-6) protein.
XX
XX Tumour necrosis factor stimulated gene 6; TSG-6; recombinant; human;
KW diagnosis; cytokine; immune disorder; autoimmune disorder; infection;
KW inflammatory disease; neurodegenerative disease; cancer; hepatitis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..19
XX FT /note= "putative signal sequence"
XX FT Protein 20..277
XX FT /note= "mature protein"
XX FT Modified-site 118..120
XX FT /note= "potential Asn-glycosylation site"
XX FT Modified-site 258..260
XX FT /note= "potential Asn-glycosylation site"
XX
XX US5846763-A.
XX
XX 08-DEC-1998.
XX
XX 13-MAY-1994; 94US-0242097.
XX
XX 13-MAY-1994; 94US-0242097.
PR 14-JAN-1991; 91US-0642312.
PR 01-MAR-1993; 93US-0024868.
XX
XX (UANY) UNIV NEW YORK STATE.
XX
XX Lee TH, Vilcek J, Wisniewski H;
XX
XX

Wed Apr 4 13:03:50 2001

us-09-466-778-11.rag

Page 2

SEQ ID 11, Data
AC NO. W84087

Seq ID 11, Database: A-Genesg-36

14-JAN-1991; 91US-0642312.
(OVNY) UNIV NEW YORK STATE.
Lee TH, Vilcek J, Wisniewski HG;
WPI; 1992-284330/34.
N-PSDB; Q27190.
Tumour necrosis factor-induced (glyco)protein mol. TSG-6 and its antibody - for treatment and diagnosis of chronic inflammatory conditions e.g. rheumatoid arthritis, infections, sepsis and cancer
Claim 1; Fig 3; 122pp; English.
The sequence is that of tumour necrosis factor (TNF) induced glycoprotein TSG-6. TSG-6 is induced in connective tissue by TNF and interleukin-7, and interferes with cell adhesion to their cells or the extracellular matrix. It and antibodies specific for it are useful in diagnosis (including identifying susceptibility) and treatment of various disorders associated with cytokine (in)activation e.g. chronic inflammation (esp. where associated with proteoglycan breakdown, such as rheumatoid arthritis) cancer (esp. development of metastases) and infections (esp. those caused by Gram negative bacteria).
Sequence 277 AA;

	Query Match	14.1%	Score 262;	DB 13;	Length 277;
	Best Local Similarity	43.3%	Pred. No. 1.6e-19;		
	Matches 4;	Conservative 16;	Mismatches 13;	Indels 0;	Gaps 0;
52	DTTVGVHLRSLGQYKLTETKAREACANEAATMAYTNQLSYQKAKYHLCISAGWLETR	111			
32	eraagvvhrearggkyklyaeakvacefegghlatkyglearklfgfhvcaagvmakgr	91			
112	VAYPTAFASQNCGGVGVYDYGPPRNKSEMDVDFCYRKMDVNC	155			
92	vggplvkgpncpcgkfgtgidlydglmrserwdaycynphakes	135			

2	RESULT	
WI3654	WI3654	
WD	WI3654 standard; protein; 277 AA.	
XX		
AC	WI3654;	
XX		
DT	07-NOV-1997 (first entry)	
XX		
DE	Tumour necrosis factor-stimulated gene-6 protein.	
XX		
XX	TSG-6; human; tumour necrosis factor-stimulated gene-6 protein; treat;	
XX	inflammatory disease; cancer-related pathology; inhibit; metastasis;	
XX	invasive tumour growth; rheumatoid arthritis; cachexia; atherosclerosis;	
XX	systemic lupus erythematosus; rational drug design.	
XX		
DS	Homo sapiens.	
XX		
PN	WC9704075-AL.	
XX		
PD	06-FEB-1997.	
XX		
PF	19-JUL-1996; 96WO-US11995.	
XX		
PP	20-JUL-1995; 95US-0001311.	
XX		
PPA	(UJNY) UNIV NEW YORK STATE.	
XX		
PI	Cronstein BN, Vilcek J, Wisniewski H;	
XX		
DR	WPI; 1997#132619/12.	
XX		

New TSG-6 compositions to treat inflammatory diseases and cancer -
 can be complexed with inter-alpha-inhibitor, to treat auto-immune
 disease, cancer-related pathology etc. can also be administered as
 TSG-6 DNA
 Claim 2, Page 58-59; 76pp; English.
 This protein is the human tumour necrosis factor-stimulated gene-6
 protein (TSG-6). It can be used in a claimed pharmaceutical composition
 for treating inflammatory diseases and disorders or cancer-related
 pathologies. The composition or DNA encoding TSG-6 may be used to inhibit
 tumour metastasis, invasive tumour growth or cancer-related pathologies.
 They may also be used in the treatment of inflammatory conditions
 including rheumatoid arthritis, systemic lupus erythematosus, and
 infections such as catheixia, atherosclerosis etc. TSG-6 can also be used
 to design drugs to treat these disorders.
 Sequence 277 AA;

	Query Match	14.1%	Score 262;	DB 18;	Length 277;	
	Best Local Similarity	43.3%;	Pred; No.	1.6e-19;		
	Matches 4;	Conservative 16;	Mismatches 43;	Indels 0;	Gaps. 0;	
OQY	52 DTTGVPHLSPLGQYLKTPDKARECANPAMATMYNOLSXXQAKYHLCASAGWLETGR	111	:	:::::	: :	:
Ddb	32 eraagvvhreargkyklytaeaakaycefeeghnltqylearkigfhvcgaawmakaqr	91	:	:	:	:
OQY	112 VAYPTAFASCSCGGVGVIYDGPENPKSEMDVCFPRMKDYNC	155	:	:	:	:
Ddb	92 vgyplvkpgnccgfgtkgiidvgirlnrsrwdacynpahacek	135	:	:	:	:

Key	Location/Qualifiers
Peptide	1..19
Protein	/note= "putative signal sequence"
Modified-site	20..277
Modified-site	/note= "mature protein"
Modified-site	118..120
Modified-site	/note= "potential Asn-glycosylation site"
Modified-site	258..260
Modified-site	/note= "potential Asn-glycosylation site"
US5846763-A.	
08-DEC-1998.	
13-MAY-1994;	94US-0242097.
13-MAY-1994;	94US-0242097.
14-JAN-1991;	91US-0642312.
01-MAR-1993;	93US-0024868.
(JYNY) UNIV NEW YORK STATE.	
Lee TH, Vilcek J, Wisniewski H:	

File Copy

WPI; 1999-059056/05.

WPI; 1999-05905
N-PSDB; V71778.

DNA encoding tumour necrosis factor stimulated gene 6 protein - useful for producing recombinant protein for diagnosis and therapy of disease e.g. immune disorders or cancer

Claim 2; Fig 3A-C; 65pp; English.

This represents a human tumour necrosis factor stimulated gene 6 (TSG-6) protein. A prokaryotic host cell transformed, or a eukaryotic host cell transfected with expression vectors containing the TSG-6 nucleic acid are used to produce recombinant human TSG-6 protein. TSG proteins, peptide fragments of TSG proteins or antibodies to TSG proteins are useful for diagnosis or therapy of diseases mediated by cytokine activity or inactivity, such as immune and autoimmune disorders, infections, inflammatory diseases, neurodegenerative diseases, cancer and alcohol-induced hepatitis.

Sequence 277 AA;

Every Match 14.1%; Score 262; DB 20; Length 277;
 t Local Similarity 43.3%; Pred. No. 1.6e-19;
 ches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;

52 DTTGVFHLRSP LGQYKLTFDKAREACANEAAATMATYNQLSYXQKAKYHLC SAGWLETGR 111

```

32 eraagvyhrearsgkyklt,yaeakavcefegghlatyqleaaarkigfhwcaagwmakgr 91

```

112 VAYPTAFASQNCGSVVGIVDYGPRPNKSEMWDVFCYRMKDVCN 155

92 vgyplvkpgpnccgfgktgldygirlnrserwdacynphakec 135

4
T

1
V13381 standard: Protein: 360 AA.

Y19387:

25-JUN-1999 (first entry)

Amino acid sequence of protein PRO271.

Secreted protein; transmembrane protein; human; enterocolitis; Zollinger-Ellison syndrome; gastrointestinal ulceration; congenital microvillus atrophy; skin disease; cell growth; abnormal keratinocyte differentiation; psoriasis; epithelial cancer; Parkinson's disease; Alzheimer's disease; ALS; neuropathy; fibromodulin; decmal scarring; Usher Syndrome; Atrophia areata; anti-thrombotic; wound healing; tissue repair.

Homo sapiens.

WO9914328-A2.

25-MAP-1990

SECRET

16-SEP-1998;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

25-NOV-1997;
17-SEP-1997;

17-SEP-1997;
17-SEP-1997;

17-SEP-1997;

17-SEP-1997;

117-SEP-1997;

17-SEP-1997;
17-SEP-1997;

18-SEP-1997:

18-SEP-1997;

15-OCT-1997;

17-OCT-1997;

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 seconds
(without alignments)
1505.071 Million cell updates/sec

Title: US-09-466-778-11
Perfect score: 1857
Sequence: 1 MTGPGKHCKESHVVDGL.....ALAAYSYFRINKTIGFXHF 353

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL15:*
- 1: sp.archaea:*
 - 2: sp.bacteria:*
 - 3: sp.fungi:*
 - 4: sp.human:*
 - 5: sp.invertebrate:*
 - 6: sp.mammal:*
 - 7: sp.mhc:*
 - 8: sp.organella:*
 - 9: sp.phage:*
 - 10: sp.plant:*
 - 11: sp.rodent:*
 - 12: sp.virus:*
 - 13: sp.vertebrate:*
 - 14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1757	94.6	897	4 Q9NRY3	Q9nry3 homo sapien
2	1468	79.1	1069	4 Q9UP98	Q9up98 homo sapien
3	662	35.6	2212	4 Q93072	Q93072 homo sapien
4	645.5	34.8	2570	4 Q9NY15	Q9ny15 homo sapien
5	264	14.2	275	11 Q08859	O08859 mus musculus
6	177	9.5	2109	13 P79787	P79787 gallus gall
7	169.5	9.1	355	11 Q921X7	Q921x7 mus musculus
8	162	8.7	1321	4 Q14594	Q14594 homo sapien
9	161.5	8.7	655	11 Q08564	Q08564 rattus norv
10	161.5	8.7	1290	13 Q9W6E1	Q9w6e1 gallus gall
11	160.5	8.6	2394	6 Q77610	Q77610 bos taurus
12	153	8.2	192	6 Q02817	Q02817 oryctolagus
13	153	8.2	656	6 Q77612	Q77612 bos taurus
14	151.5	8.2	1643	6 Q77611	Q77611 bos taurus
15	151.5	8.2	3381	6 Q77609	Q77609 bos taurus
16	126.5	6.8	103	6 Q9T83	Q9t83 sus scrofa
17	120.5	6.5	103	6 Q46380	Q46380 oryctolagus
18	110	5.9	396	13 Q9W6S4	Q9w6s4 gallus gall
19	109	5.9	302	4 Q9UNF4	Q9unf4 homo sapien

20	109	5.9	322	4 Q9Y5Y7	Q9y5y7 homo sapien
21	105	5.7	1584	5 Q93791	Q93791 caenorhabdi
22	104	5.6	3507	5 Q23587	Q23587 caenorhabdi
23	103.5	5.6	717	13 P87357	P87357 brachydanio
24	102.5	5.5	778	13 Q9IBG4	Q9ibg4 xenopus lae
25	100	5.4	816	11 Q70474	Q70474 rattus norv
26	99.5	5.4	682	4 Q9NTW2	Q9ntw2 homo sapien
27	98.5	5.4	728	13 Q90656	Q90656 gallus gall
28	98.5	5.4	802	13 Q57462	Q57462 brachydanio
29	99.5	5.4	3623	4 Q60494	Q60494 homo sapien
30	99	5.3	780	11 Q9QYX8	Q9qyx8 mus musculu
31	99	5.3	1440	5 Q20204	Q20204 caenorhabdi
32	99	5.3	2906	11 Q9WUH9	Q9wuh9 rattus norv
33	98	5.3	294	4 Q92493	Q92493 homo sapien
34	98	5.3	2531	5 Q16004	Q16004 lytechinus
35	97.5	5.3	780	11 Q08779	Q08779 rattus norv
36	97.5	5.3	1480	5 Q9V7F8	Q9v7f8 drosophila
37	97.5	5.3	1504	5 Q9XV4	Q9xv4 drosophila
38	97.5	5.3	1504	5 Q9V7F9	Q9v7f9 drosophila
39	96	5.2	364	11 Q70509	Q70509 rattus norv
40	96	5.2	742	4 Q9UJ36	Q9uj36 homo sapien
41	95.5	5.1	364	6 Q97569	Q97569 ceratotheri
42	95.5	5.1	721	13 Q91902	Q91902 xenopus
43	95	5.1	261	10 Q24530	Q24530 vitis vinif
44	94.5	5.1	1328	13 P79754	P79754 fugu rubrip
45	94	5.1	637	10 Q9SF50	Q9sf50 arabidopsis

ALIGNMENTS

RESULT 1

Q9NRY3 ID Q9NRY3 PRELIMINARY; PRT; 897 AA.
AC Q9NRY3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE CD44-LIKE PRECURSOR FELL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tao Q., Zhang W., Cao X.;
RT "Molecular cloning and characterization of human FELL sharing homology with CD44".
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF160476; AAF62398.1; .
SQ SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match 94.6%; Score 1757; DB 4; Length 897;
Best Local Similarity 93.8%; Pred. No. 8.6e-162;
Matches 331; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY	1	MTGPGKHCKESHVVDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTGVGFHL	60
DB	489	MTGPGKHCKESHVVDGLNCEPEQLPDRCLQDNGQCHADAKCVDLHFQDTTGVGFHL	548
QY	61	RSPLQYKLTDFDKAREACANEAATMATYNQLSYQKAKYHLCAGWLETGRVAYPTAFAS	120
DB	549	RSPLQYKLTDFDKAREACANEAATMATYNQLSYQKAKYHLCAGWLETGRVAYPTAFAS	608
QY	121	QNCGSGVGVVDYGRPNKSEMDVFCYRMKDVNCTKVGKVGDFSGFSGNLLQVLMSEF	180
DB	609	QNCGSGVGVVDYGRPNKSEMDVFCYRMKDVNCTKVGKVGDFSGFSGNLLQVLMSEF	668
QY	181	SLTNELTEVLAYSNSARGRAFLHLEHTDLSIRGTLPQNSGLGENETLSGRDIEHHLAN	240
DB	669	SLTNELTEVLAYSNSARGRAFLHLEHTDLSIRGTLPQNSGLGENETLSGRDIEHHLAN	728

File Copy


```

QY 241 VSMFYNDLVNLTQTRGLSKLLITRDQPLHPTETRCVGDGHTLEWDICASNGITHVI 300
DB 729 VSMFYNDLVNLTQTRVSGSKLLITASQDPLQPTETRFVDGRAILQWDFASNGIIRVI 788
QY 301 SRXLKAPPAPVTLHTXGLGIXFIIXIILVTGAVALAAYSFRINRRTIGFQHF 353
DB 789 SRPLKAPPAPVTLHTXGLGIXFIIXIILVTGAVALAAYSFRINRRTIGFQHF 841
DE DKEZP434E0321.
GN DKEZP434E0321.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId-9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133021; CAB61358.1;
DR HSSP: P98066; ITSG.
DR INTERPRO: IPR000538;
DR INTERPRO: IPR000561;
DR INTERPRO: IPR000782;
DR INTERPRO: IPR002049;
DR PFAM: PF00008; EGF_8;
DR PFAM: PF00193; Xlink; 1;
DR PRINTS: PR00011; EGF_LIN;
DR PRINTS: PR001265; LINKMODULE.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 5.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1069 AA; 115666 MW; 0480960932164D63 CRC64;

Query Match 79.1%; Score 1468; DB 4; Length 1069;
Best Local Similarity 80.5%; Pred. No. 1.2e-133;
Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 MTGPKHKCKESHVYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHL 60
DB 697 MTGPKHKCKESHVYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHL 756
QY 61 RSLPGQYKLTDPKAREACANEAATMATYNQLSYQAKYHLCAGWLETGRVAYPTAFAS 120
DB 757 RSLPGQYKLTDPKAREACANEAATMATYNQLSYQAKYHLCAGWLETGRVAYPTAFAS 816
QY 121 QNCGSGVGVYDYGPRPNKSEMDVFCYRMKDVNCTKXKVGVDGFSYSGNLLQVLMSP 180
DB 817 QNCGSGVGVYDYGPRPNKSEMDVFCYRMKDVNCTKXKVGVDGFSYSGNLLQVLMSP 860
QY 181 SLTNFLTEVLAYSNSARGRAFLHLEHLDLSIRGTLFQPNQSGLGENETLSGRDIEHHLA 240
DB 861 SLTNFLTEVLAYSNSARGRAFLHLEHLDLSIRGTLFQPNQSGLGENETLSGRDIEHHLA 900
QY 241 VSMFYNDLVNLTQTRGLSKLLITRDQPLHPTETRCVGDGHTLEWDICASNGITHVI 300
DB 901 VSMFYNDLVNLTQTRGLSKLLITASQDPLQPTETRFVDGRAILQWDFASNGIIRVI 960
QY 301 SRXLKAPPAPVTLHTXGLGIXFIIXIILVTGAVALAAYSFRINRRTIGFQHF 353
DB 961 SRPLKAPPAPVTLHTXGLGIXFIIXIILVTGAVALAAYSFRINRRTIGFQHF 1013

```

```

RESULT 3
QY 93072 PRELIMINARY; PRT; 2212 AA.
ID Q93072;
AC Q93072;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MTEOBLAST KIAA0246 PROTEIN (FRAGMENT).
GN KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId-9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE-9719144; PubMed-9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.
RL DNA Res. 3:321-329(1996).
DR EMBL: D87433; BAA1337.1;
DR HSSP: P98066; ITSG.
DR INTERPRO: IPR000538;
DR INTERPRO: IPR000561;
DR INTERPRO: IPR000782;
DR INTERPRO: IPR001128;
DR PFAM: PF00008; EGF_13;
DR PFAM: PF00193; Xlink; 1;
DR PRINTS: PR01265; LINKMODULE.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 13.
DR PROSITE: PS01241; LINK; UNKNOWN_1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
KW Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 2212 AA; 237451 MW; 4A95480504129134 CRC64;

```

```

Query Match 35.6%; Score 662; DB 4; Length 2212;
Best Local Similarity 40.2%; Pred. No. 4.9e-55;
Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGFGKHKCKESHVYVGDGLNCEPEQLPIDRCLQDNGQCHADAKCYDLHFQDITVGVFHL 60
DB 1796 TGLNTRCCHAGYVGDGLQCLESEPPVDRCLQGPCHSDAMCTDLHFQEKRAQVFL 1855
QY 61 RSLPGQYKLTDPKAREACANEAATMATYNQLSYQAKYHLCAGWLETGRVAYPTAFAS 120
DB 1856 QATSGPYGLNFEAEAEACAEAGAVLASEPQLSAAQQLGFHLCLMGLANGSTAHPPVFPV 1915
QY 121 QNCGSGVGVYDYGPRPNKSEMDVFCYRMKDVNCTKXKVGVDGFSYSGNLLQVLMSP 179
DB 1916 ADCNGRGVIGLGNKLSERDAICFRQVDVACRNGFVGDICTCKGLDVLAAI 1975
QY 180 SLTNFLTEVLAYSNSARGRAFLHLEHLDLSIRGTLFQPNQSGLGENETLSGRDIEHHLA 239
DB 1976 ANESTFVGLLSYANATQGLDFDLDELTYKTFLFPVNEGFVDNLTSLGPNLEHAS 2035
QY 240 VSMFYNDLVNLTQTRGLSKLLITDR---QDPLHPTETRCVGDGHTLEWDICASNGI 296
DB 2036 NATLLSAN-ASQCKLLPAHSGLSLIISDAGDPNNSWAPAGPTVTVSRIVWDIMAFNGI 2094
QY 297 TVHSRKLKAPPAPVTLHTXGLGIXFIIXIILVTGAV---ALAAYSYFRINRRTIGF 350
DB 2095 IHALASPLAPQPPAVLAPEAPPVAGVAVLAAGALLGLVAGALYLRARGKPMGF 2150

```

